AGARD-CP-453

AGARD-CP-453

AD-A215 917

AGARD

ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92200 NEUILLY SUR SEINE FRANCE

AGARD CONFERENCE PROCEEDINGS No.453

Operational Decision Aids for Exploiting or Mitigating Electromagnetic Propagation Effects

NORTH ATLANTIC TREATY ORGANIZATION

DISTRIBUTION AND AVAILABILITY
ON BACK COVER

AGARD-CP-453

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE L'ATLANTIQUE NORD)

AGARD Conference Proceedings No.453

OPERATIONAL DECISION AIDS FOR EXPLOITING OR MITIGATING ELECTROMAGNETIC PROPAGATION EFFECTS

Papers presented at the Electromagnetic Wave Propagation Panel Symposium held in San Diego, California, USA, 15—19 May 1989.

THE MISSION OF AGARD

According to its Charter, the mission of AGARD is to bring together the leading personalities of the NATO nations in the fields of science and technology relating to aerospace for the following purposes:

— Recommending effective ways for the member nations to use their research and development capabilities for the common benefit of the NATO community;

— Providing scientific and technical advice and assistance to the Military Committee in the field of aerospace research and development (with particular regard to its military application);

— Continuously stimulating advances in the aerospace sciences relevant to strengthening the common defence posture;

— Improving the co-operation among member nations in aerospace research and development;

— Exchange of scientific and technical information;

— Providing assistance to member nations for the purpose of increasing their scientific and technical potential;

— Rendering scientific and technical assistance, as requested, to other NATO bodies and to member nations in connection with research and development problems in the aerospace field.

The highest authority within AGARD is the National Delegates Board consisting of officially appointed senior representatives from each member nation. The mission of AGARD is carried out through the Panels which are composed of experts appointed by the National Delegates, the Consultant and Exchange Programme and the Aerospace Applications Studies Programme. The results of AGARD work are reported to the member nations and the NATO Authorities through the AGARD series of publications of which this is one.

Participation in AGARD activities is by invitation only and is normally limited to citizens of the NATO nations.

The content of this publication has been reproduced
directly from material supplied by AGARD or the authors.

Published September 1989

ISBN 92-835-0524-7

Printed by Specialised Printing Services Limited
40 Chigwell Lane, Loughton, Essex IG10 3TZ

THEME

Systems relying on electromagnetic, including electro-optic wave propagation in the earth's environment are subject to propagation anomalies. Proper assessment of such anomalies and subsequent exploitation or mitigation of their effects often require complex computations. The availability of small powerful computers has made near real-time calculation of propagation effects on systems performance possible and affordable. For example, the Integrated Refractive Effects Prediction System (IREPS) and the HF assessment system PROPHET have been used successfully throughout the military community for some ten years. During that time a number of operational decision aids (ODAs) were developed like optimum flight altitudes for attack, surveillance and ESM aircraft; intercept vulnerability; and ODAs for command and control. Since operational decision aids require appropriate propagation models and accurate environmental inputs, careful attention must be paid to propagation modelling, direct and remote sensing techniques and environmental forecasting techniques.

The topics covered include:

1. General aspects;
2. Propagation modelling and validation;
3. Decision aids for tropospheric radio propagation;
4. Decision aids for ionospheric radio propagation;
5. Decision aids for electro-optical propagation;
6. Remote and direct sensing techniques;
7. Forecasting of environmental parameters;
8. Operational systems impact.

Exploitation or mitigation of environmental effects rank equal in importance with weapons systems. The rapidly changing propagation environment throughout NATO's area of concern requires a rapid operational assessment capability.

* * *

Les systèmes qui font appel à la propagation des ondes électromagnétiques et électro-optiques en milieu terrestre sont sujets à des anomalies de propagation.

Toute évaluation correcte de telles anomalies ainsi que l'éventuelle exploitation ou atténuation de leurs effets passent par des calculs qui sont souvent complexes. L'arrivée sur le marché de petits ordinateurs performants a permis le calcul des effets de propagation en quasi-temps réel pour un coût abordable. A titre d'exemple, le Système Intégré de Prévision des Effets Réfringents (IREPS) et le système de prévision RF "PROPHET" sont utilisés avec succès par la communauté militaire depuis dix ans. Au cours de cette même période un certain nombre d'aides à la prise de décisions opérationelles (ODA) ont été développées dans certains domaines tels que les altitudes de vol optimales préalables au passage d'attaque, la surveillance et les avions ESM, la vulnérabilité à l'interception et les ODA de commandement et contrôle. Les aides à la prise de décisions opérationnelles sont tributaires de modèles de propagation appropriés et de données précise sur le milieu ambiant. Il importe donc d'accorder une attention toute particulière à la modèlisation de la propagation, aux techniques de détection et de télédétection et aux méthodes de prévision du milieu ambiant.

Parmi les sujets développés, on distingue:

1. Aspects généraux;
2. La modélisation de la propagation et sa validation;
3. Les aides à la prise de décisions pour la propagation RF troposphérique;
4. Les aides à la prise de décisions pour la propagation RF ionosphérique;
5. Les aides à la prise de décision pour la propagation électro-optique;
6. Les techniques de détection et de télédétection;
7. La prévision des paramètres du milieu ambiant;
8. L'impact des systèmes opérationnels.

L'exploitation ou l'attenuation des effets du milieu ambiant sont d'une importance égale à celle des systèmes d'armes eux-mêmes. L'environnement de propagation est en évolution constante et les pays membres de l'OTAN doivent donc disposer de moyens d'analyse de leurs capacités opérationnelles, d'une mise en œuvre [illegible].

ELECTROMAGNETIC WAVE PROPAGATION PANEL

Chairman: Prof. C.Goutelard
Directeur L.E.T.T.I.
Université Paris-Sud
9, Av. de la Division Leclerc
94230 Cachan
France

Deputy Chairman: Ir. H.Vissinga
FEL-TNO
PO Box 96864
2509 JG The Hague
Netherlands

TECHNICAL PROGRAMME COMMITTEE

Co-Chairmen

Dr H.J.Albrecht
FGAN
Neuenahrer Strasse 20
D-5307 Wachtberg-Werthhoven
Germany

Dr J.Richter
Head Ocean and Atmospheric Sc. Division, Code 54
NOSC
San Diego, CA 92152-5000
United States

Members

Dr T.K.Fitzsimons
ARFA
NATO HQs
B1108 Bruxelles
Belgium

Lt Col. G.C.Kunz
SHAPE
CISD
7010 Shape
Belgium

Lt Col. D.Scholz
FAFCO
Tomburg-Kaserne
Münstereifeler Str. 75
D-5308 Rheinbach
Germany

Dr P.F.Twitchell
Code 1214
Office of the Chief of Naval Research
800 North Quincey Street
Arlington, VA 22217
United States

Col. T.Cress
Military Assist. for Envir. Sc.
R&AT/ELS
Office of the Dir. of Defense
Research and Engineering
Room 3D129, Pentagon
Washington, DC 20301
United States

Capt. J.J.Jensen
Program Manager for Envir. Syst.
SPAWAR 141
Space & Naval Warfare Syst. Com.
Washington, DC 20363-5100
United States

Dr R.B.Gomez
Associate Technical Director
Engineer Topographic Lab.
For Belvoir, VA 22060-5546
United States

AGARD EPP EXECUTIVE

Lt Col. P.A.Brunelli

From Europe
AGARD
7, rue Ancelle
92200 Neuilly sur Seine
France

From the USA and Canada
NATO/AGARD
APO NY 09777

FOREWORD

The operational use of decision aids dealing with EM/EO atmospheric propagation effects is the final link in a long chain of research and development endeavours. Often lengthy and intense basic research investigations, both theoretical and experimental, provide a basic understanding and description of the phenomena involved. When this knowledge is combined with an appreciation of military operational problems, appropriate assessment tools and decision aids can be devised. The most successful decision aids are those which are developed in close interaction with the user. It was the purpose of the Symposium "Operational Decision Aids for Exploiting or Mitigating Electromagnetic Propagation Effects" to bring researchers and operational users together to promote this all-important interaction. The meeting covered atmospheric propagation problems throughout the electromagnetic spectrum from long wave propagation through optical frequencies and included environmental forecasting topics and man-machine interface issues.

The symposium was held in San Diego, CA from 15—19 May 1989 and included six different sessions. An overview session included several review papers. It was followed by a session on decision aids and systems impact. The next three sessions covered tropospheric radio, ionospheric radio and electro-optical propagation topics. The final session was concerned with assessment and forecasting. All together, 50 papers were presented during the week-long meeting.

Gratefully acknowledged are the cooperation and assistance of the Technical Programme Committee, Colonel Cress, Dr Fitzsimons, Dr Gomez, Captain Jensen, Lieutenant Colonel Kunz, Lieutenant Colonel Scholz and Dr Twitchell.

Appreciation is furthermore expressed to the AGARD staff led by Lieutenant Colonel Brunelli and the local coordination group led by Mrs Nancy Vorce for the excellent organization of the meeting.

H.J.Albrecht
J.H.Richter

CONTENTS

*Printed in classified publication CP453 (Supplement).

*Printed in classified publication CP453 (Supplement).

THE IMPORTANCE OF ENVIRONMENTAL DATA

Gerald Cann and Robert P. Porter
Members of the
Naval Research Advisory Committee
Washington, DC 20350 USA

SUMMARY

A panel of experts conducted a top level analysis of the significance of quantitative knowledge of environmental parameters to naval weapons and naval warfare. The panel concluded that the impact of the environment needs to be considered more thoroughly during the research, development and acquisition process. This requires a central environmental top level requirement or master plan and a more formalized connection between the organizations involved in various aspects of environmental science.

INTRODUCTION

The Navy Research Advisory Committee (NRAC) is the U.S. Navy's senior scientific advisory group to the Secretary of the Navy, Chief of Naval Operations, Commandant of Marine Corps and Chief of Naval Research. It is dedicated to providing independent and objective analyses in areas of science, research, technology and development. In the summer of 1988, NRAC convened a panel of experts to conduct a top level review of "Importance of Environmental Data to Naval Weapons and Warfare." The panel looked at a wide range of environmental factors that influence Naval operations. Environmental knowledge and the proper consideration of its effects are essential throughout the entire acquisition process from basic research through development, test, evaluation and production. After deployment, operations and tactics must take environmental conditions into account and either mitigate or exploit their effects. The panel stressed the importance of satellite remote sensing and the need for national and international cooperative efforts in this area. It made a number of other recommendations including the need for more formalized consideration of environmental factors in the research, development and acquisition process.

BACKGROUND AND OBJECTIVE

It is widely recognized that Anti-Submarine Warfare (ASW) techniques and equipment are dependent upon acoustic properties. Therefore, it has been postulated that precise knowledge of the acoustic properties of specific bodies of water could significantly enhance the effectiveness of our ASW forces. Similarly, since the atmospheric environment controls the propagation of Radio Frequency (RF) energy, it has been postulated that naval forces can fight more effectively if they can reliably estimate these factors. Laser and other advanced weapons and sensors are equally influenced by the natural environment. The general objective of the NRAC 1988 Summer Study Panel on the Importance of Environmental Data was to determine the significance of quantitative knowledge of environmental parameters to naval weapons and warfare. Specific tasking included the questions: Which weapon systems are affected; how much performance can be gained; how precise must the measurements be; and, what type of sensors/techniques are needed?

The panel first attempted to identify the most critical issues, and quickly homed in on perceived deficiencies in the Navy's organizational structure and procedures related to the impact of the environment, particularly in the Research, Development and Acquisition (RDA) process. This became a major focus of the study.

There is no question that all naval weapon systems are susceptible to performance degradation by variable and uncompensated environmental features. There is no doubt that real-world weapon performance can be enhanced by improved knowledge and exploitation of the environment. The real issue is one of priorities. The panel addressed this issue by selecting and presenting examples which have the following characteristics:

1. The systems represented are of paramount importance to the Navy;
2. The benefit from adequate consideration of environmental effects may be the difference between mission success and failure;
3. The need for certain specific environmental measurements is clearly indicated by the examples.

Establishment of the quantitative relationships of system performance versus environmental characteristics, as expressed in the terms of reference, should be an on-going process in the developer and user communities of the Navy. It was beyond the capacity of this NRAC panel, in terms of available time and resources, to reach beyond the identification of important parameters, the needs for environmental information, and the improvement in policies and procedures required to integrate them efficiently into the acquisition and fleet operations processes.

Some of the derivative issues addressed included the application of stealth technology to threat platforms and missiles which causes their signatures to "sink" deeper into the environment and tends to frustrate our target detection systems. This evolution must drive our effort to improve the understanding and exploitation of environmental characteristics, and has significantly determined the direction of this study.

The panel did not attempt to thoroughly examine the environmental requirements needed to support future sensor systems that may emerge as threat signatures are suppressed. For example, low frequency magnetic, spatial gradient magnetic anomaly, and bioluminescent sensors were not addressed. It is clear that some of these possible future sensors (e.g., Electro-Optical (EO)) may need more environmental data support than present systems require.

The panel recognized, at the outset, the importance of improving our ability to assess environmental effects on future weapons throughout the acquisition life cycle--from initial concept development through deployment--including impacts on training, tactics and fleet operations.

IMPORTANCE OF ENVIRONMENTAL KNOWLEDGE

The next war will not be a war of attrition at sea since assets are bounded. The number of ships, aircraft and precision guided weapons are well defined. Therefore, it is extremely important to maximize the success of the first engagement.

Currently we exploit first order environmental effects. The Navy has a large network to collect and disseminate weather and oceanographic data. At this point in time the easy things have been done. Detailed data bases to support future requirements are not available.

The advent of low observables requires tightening all performance factors. Quiet submarines and low radar cross-section aircraft/missiles will make fine-grained environmental data very important. As signal to noise decreases for current systems, it will be necessary to factor the environment into concepts, planning and tactics to defeat these targets.

Low intensity conflicts/crises tend to exacerbate the need for environmental data. Over the last forty years, many of the limited war crises/conflicts have been in placed where U.S. forces have had to operate in coastal waters where acoustic conditions are bad, local bathymetry is unknown, and there are complex RF/Infrared (IR)/EO sea/land interface conditions. As obvious example of this is the recent Persian Gulf experience.

From the perspective of this panel, there appears to be a high payoff from exploiting environmental effects. However, these effects are often quite subtle, and must generally be considered on a case-by-case basis for each weapon system individually. A detailed understanding of such specific issues as propagation paths and optimum frequencies could result in significantly improved acoustic, IR, EO and Electromagnetic (EM) system design and operation.

Comprehensive assessment of requirement/concepts/hardware/tactics/training with respect to physical limits imposed by the environment needs to be a high priority consideration by decision makers prior to major financial commitment to new systems. This is not generally occurring in the RDA process, and the panel identified this as a deficiency in the system.

The Navy's requirement for environmental information is global in nature. It not only involves the ocean environment from the tropics to the poles, but also the coastal and land environments. Some of the areas are either not accessible or access to them may be denied.

Environmental knowledge and the proper consideration of its effects are essential throughout the entire acquisition process from basic research through development, test, evaluation and production. After deployment, operations and tactics must take environmental conditions into account and either mitigate or exploit their effects.

There are many environmental factors that influence naval operations. Most obvious are weather effects and properties of the ocean, both surface and subsurface. Less obvious, but often of crucial importance, are properties of the sea floor and soil characteristics of the adjacent land. Bottom topography of the ocean floor is critical to submarine operations and sound propagation; in shallow water it also affects mining and amphibious warfare. Soil conditions on the beach and the adjacent land may have a significant impact on landing operations. For example, unexpected sand and dust storms may not only reduce visibility for the human eye and EO sensors, but may also render machinery, such as helicopter and tank engines, inoperable. The often highly variable structure of the atmosphere profoundly affects EM (including EO) propagation. Similarly, underwater acoustic propagation is critically dependent on temperature and salinity changes in the ocean.

There must be a continuous interaction between the collectors and analysts of environmental data and the users in the research through operations process described above.

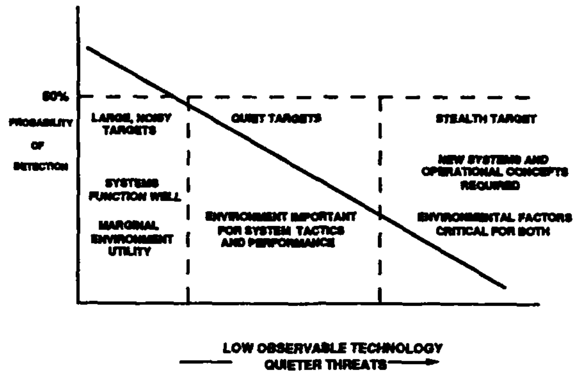


Figure 1. Importance of Environmental Data

The impact of environmental data on improved performance of a system depends on the ratio of target signal to background fluctuations as perceived by the system and illustrated in figure 1.

When the target signal is extremely strong and stands out above the background, the system functions well and knowledge of the environment provides only marginal system performance improvements.

For the case where the target signal is quiet and is often obscured by the ambient background fluctuations, it is very important to exploit the environmental data to improve system performance, platform development strategy, and tactics during engagement. Proper use of environmental data can make the difference between operational failure and success.

When the target becomes extremely stealthy, and the target signal is masked by the background fluctuations, new systems and new operational concepts are required. Environmental factors will be critical in the design and development of improved new systems and tactics which maximize system performance.

Low observable technology is being used to produce quieter, more formidable threats. The techniques of employing environmental data to counter the quieting and successfully overcome stealthier threats are threefold.

In figure 2, success most often occurs when the system performance curve is above 50 percent probability of detection. Failure most often occurs when the system performance falls below 50 percent probability of detection.

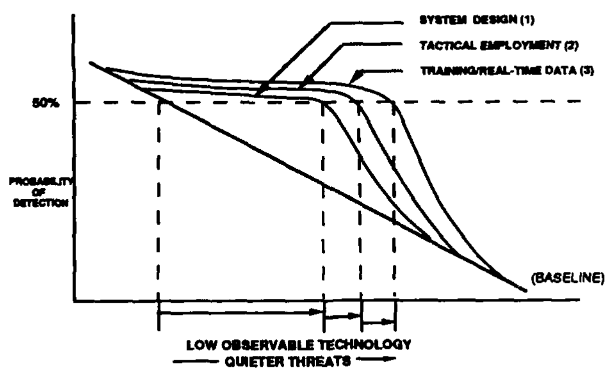


Figure 2. Value of Environmental Data

The baseline curve in figure 2 represents present system performance capability which begins to fail as quieter threats are engaged. The first usage of environmental data is to help design improved systems which better exploit the environment. The system design improvement which gives the operator the right tool is illustrated by curve (1). Note that much quieter threats can be detected with the improved tool.

Tactics of force employment and asset positioning which take advantage of accurate environmental data to put the platforms and sensors in the proper place with the correct orientation to achieve successful engagement represent the second usage of environmental data. When combined with improved systems, this technique yields the system performance curve (2).

The third technique involves environmental training and real-time environmental data to operationally select the proper sensor modes and weapon presets. When an expert operator, informed of the up-to-date situation, uses improved tools at the right place and time in the most effective fashion, then curve (3) results.

As the threat employs even better stealth techniques, the cycle begins again because new system and operational concepts will be required. We must employ environmental data at the early stages of system design as well as throughout the life cycle and operation of the system. Further, we must understand that the most effective use of environmental data recognizes the involving of system parameters, threat characteristics, tactics, training and the environment itself.

USE OF ENVIRONMENTAL DATA

Environmental data are obtained in a variety of activities as illustrated in figure 3. For example, the technology base community develops a basic understanding of oceanographic and atmospheric phenomena. Other technology base investigators may develop the understanding of the physics of acoustic and EM wave propagation. This information is combined into predictive models for weather, oceanographic processes and wave propagation. Once the models have been developed, they may be used, to integrate inputs from a great number of variety of sensors to produce products suitable for delivery to the operations community.

- DIFFERENT ACTIVITIES USE ENVIRONMENTAL DATA IN DIFFERENT WAYS
- NO ONE ACTIVITY PROVIDES COMPLETELY ADEQUATE INFORMATION TO OTHER ACTIVITIES
- CONNECTIVITY BETWEEN ACTIVITIES IS INADEQUATE
- EACH ACTIVITY NEEDS APPROPRIATE SPECIALISTS TO EFFECTIVELY INTERACT WITH OTHER ACTIVITIES
- FLEET OPERATORS NEED BETTER PRODUCTS, DELIVERY, AND TRAINING
  - ACCURATE, TIMELY TACTICAL DECISION AIDS
  - EMBEDDED INTO WEAPON SYSTEM IF POSSIBLE

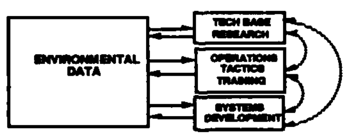


Figure 3. Use of Environmental Data

In the operations community, the environmental data may be used to assess and optimize weapon system and sensor performance, assess vulnerability, decide tactics, or simply minimize fuel consumption. The timeliness and form in which the environmental data products are provided is extremely important.

Systems development activities must consider the impact of a variable environment on system performance in the design phase and require environmental inputs during planning and execution of tests and demonstrations. System performance models, validated by tests, are a valuable tool in the design optimization process and provide focus for directed research and development.

Thus, it is clear that environmental data are required by different activities for different reasons and used in different ways. Because of this fact and the time spent on focused developments, rarely does one activity provide a product that precisely suits the needs of another activity. Therefore, each activity needs appropriate specialists and efforts that interact with the other activities to promote better connectivity and product value.

It is especially important that the information provided to the fleet operators be accurate, timely and presented in a form they can and will use. This means the raw data must be digested to the maximum extent possible, using models to produce the specific products/forecasts actually needed. Adequate communications must be provided, and on board graphical displays and tactical decision aids are necessary. The Tactical Environmental Support System (TESS) provides this link as illustrated in figure 4. TESS is a delivery system which, in its future versions (TESS(3) and beyond), will link an operator at a sophisticated computer graphics workstation on a ship in the fleet to data from local and worldwide environmental sensors, as well as to regional oceanographic centers and global computational centers.

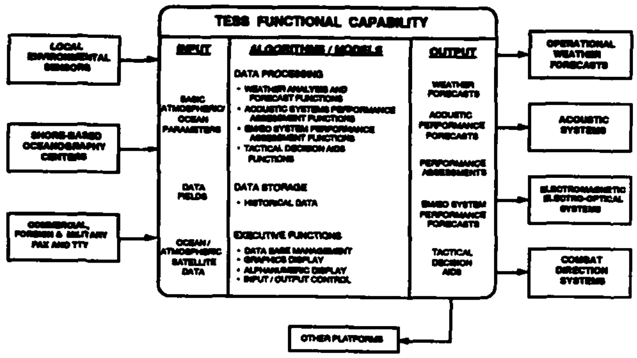


Figure 4. Tactical Environmental Support System (TESS)

TESS incorporates models to provide an on-site local predictive capability and tactical decision aids. TESS is the final link in a system that exemplifies the direction the Navy must go in providing environmental data to the fleet, and should be strongly supported.

REMOTE SENSING

Remote sensing by satellite, aircraft and earth-based sensors has been an important issue for the Navy for many years. It is widely recognized that global and synoptic coverage are applicable to naval operations and tactics, and that future systems can be optimized to take advantage of these data. The Navy Remote Ocean Sensing Satellite (NROSS) was proposed as a cooperative program with National Aeronautical and Space Administration (NASA) and National Oceanographic and Atmospheric Administration (NOAA) to support these needs. The failure to fund NROSS was partly a consequence of a poorly defined connection between the data collected and the specific requirements of naval systems. It was hard to justify NROSS on an operational basis, because the applications of the data that would have been collected were still being researched.

This panel took a close look at remote sensing from an integrated viewpoint, where some measurements were made by satellite, some by aircraft and some by earth-based systems. We focused on future environmental data requirements to meet the challenge of low observability. It is our opinion that remote sensing, especially from space-based systems, is essential to the Navy, but the cost may be too high to justify a dedicated Navy satellite. Therefore, we recommend that the Navy place a high priority on working cooperatively with other national and international agencies to obtain satellite environmental data needed for its systems, and that the Navy commit itself to supporting its own programs to integrated data collection into its operational needs.

The satellite is not the only remote sensor platform the Navy needs. Other examples include the shipborne laser radar (LIDAR) which can measure atmospheric parameters important to local weather prediction, including radar ducting. Airborne sensors can be used to measure shallow water bathymetry. Some applications require the integration of a wide variety of environmental data collected from diverse sensors. The oceanic sound speed field, critical to the efficient and effective employment of active and passive surveillance systems, ASW sonars and other fleet assets, requires satellite remote sensors, earth-based remote sensors, and in-situ ocean measurements.

Satellites and aircraft employ radar altimeters for measuring sea surface height, waveheight, wave direction and wavelength. Microwave and infrared radiometers have been well developed form measuring sea surface temperature. The microwave scatterometer can measure wind speed and direction. Synthetic aperture radars give high resolution views of the earth and can be used for mapping ice and denied coastal regions. Microwave sounders have been used to measure water vapor.

Ocean tomography is an example of an earth-based, wide-area oceanographic sensing technique which remotely maps the oceanic sound speed field. Over-the-horizon radar can measure sea state.

CONCLUSIONS AND RECOMMENDATIONS

The panel conducted a top level analysis of the significance of quantitative knowledge of environmental parameters to naval weapons and naval warfare. While it is readily apparent that all naval platforms and weapon systems are affected to some degree by the environment, the panel reviewed a number of scenarios where the impact of the environment was significant. There are combinations of weapon systems and environmental parameters where knowledge of the environment can result in order-of-magnitude improvements in system performance. However, quantifying exactly how much performance can be gained and what degree of precision of environmental measurements are required must be determined on a case-by-case basis and was beyond the scope of this panel. Adequately characterizing the marine environmental in scope of this panel. Adequately characterizing the marine environment in space and time requires the use of remote sensing techniques, both ground-based and space-based. Using remotely sensed data supported by in-situ measurements, ocean/atmosphere models will be able to accurately predict the state of the environment and weapon system performance.

As the trend towards stealth continues to drive the signatures of platforms and weapons deeper into the environment, knowledge of the environment and how it affects weapon systems continues to grow in importance. The importance and utility of environmental knowledge continues to increase until the signature is so deep in the environmental noise that is beyond detection, even in the most favorable environment. At this point, performance can be improved only through the development of new technologies and better system design. Knowledge of the environment is equally vital in system design.

The panel recommended that each new naval and marine system concept should be evaluated to determine the effect of the environment on system performance and what environmental measurements must be obtained to support the system. In addition, a top level environmental master plan should be developed and communications/interactions between the different organizations involved in various aspects of environmental science and its impact on naval operations should be connected by a formalized network.

PROPAGATION ASSESSMENT AND TACTICAL DECISION AIDS

Juergen H. Richter
Ocean and Atmospheric Sciences Division
Naval Ocean Systems Center
San Diego, CA 92152-5000 USA

SUMMARY

The evolution of real time electromagnetic/electro-optic propagation assessment systems for the US Navy is reviewed and applications of such systems to the development of Tactical Decision Aids (TDAs) are described.

INTRODUCTION

The first type of propagation assessment system described addresses propagation effects in the troposphere in the frequency range from approximately 100 MHz to 20 GHz. Figure 1 shows anomalous propagation effects when a surface-based duct is present. This situation may occur when a warm, dry layer of air lies above cooler, moist marine air. The electromagnetic wave fronts may be bent and propagate to ranges beyond the normal radio horizon. At the same time, areas of reduced coverage may result in blind spots or "radar holes". Height-finder radars, which use the the measured angle of arrival to calculate altitude, can be subjected to gross errors under such conditions. To provide an operational assessment capability for tropospheric propagation effects, a system called IREPS (Integrated Refractive Effects Prediction System) has been developed [1] and successfully used in the fleet for over 10 years. A number of TDAs were developed (such as optimum flight altitude for attack, jamming, and surveillance aircraft or maximum signal intercept range) [2,3] and have been used effectively under operational conditions. Recently, a companion system to IREPS has been completed called EREPS (Engineer's Refractive Effects Predi· tion System) which is specifically designed for the development community [4].

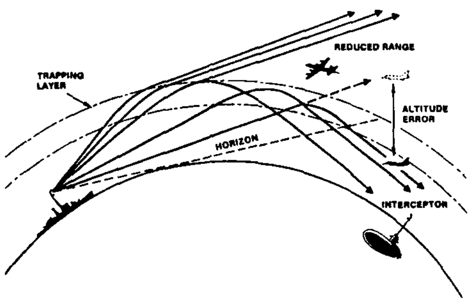


Figure 1. Propagation Anomalies caused by Refractive Layers

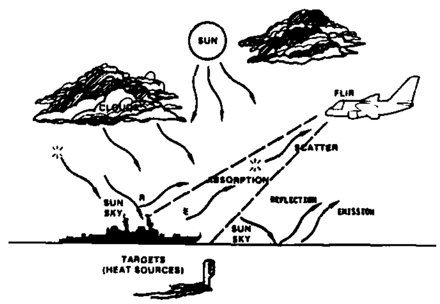


Figure 2. Electro-optical Propagation Effects in a Marine Environment

The companion to IREPS for electro-optical systems is called PREOS (Prediction of Performance and Range for Electro-optical Systems) [5] and is presently concerned with FLIR (Forward Looking Infrared) range predictions. Figure 2 illustrates several environmental effects which have to be considered if a performance prediction for a FLIR is to be made. Unlike radar range assessment where the target signature is generally independent of the environment, the FLIR signature (target-background contrast temperature) is a function of many environmental parameters. Both IREPS and PREOS are now part of the Navy's Tactical Environmental Support System (TESS) [6].

For high frequency (hf) radio propagation (2-32 MHz), a family of assessment systems was developed which is known under the name of PROPHET (for Propagation Forecasting Terminal) [7-9]. The system considers solar effects on the ionosphere (both diurnal and seasonal as well as effects caused by solar disturbances) and provides path-specific systems performance data. PROPHET and a number of TDAs for communications optimization and intercept applications have been successfully used in an operational environment for over 12 years.

TROPOSPHERIC RADIO PROPAGATION ASSESSMENT AND TDAS.

Interest in marine refractive effects arose in the 1940s when radars and communications equipment became available which operated at frequencies above approximately 100 MHz. Intense scientific efforts produced good initial understanding of the physical processes involved, some limited mathematical modeling capability, refractivity sensing techniques, and, very importantly, some excellent simultaneous radio propagation and meteorological support measurements which are still used today in model development and validation studies [10-12]. During the 1950s, the evaporation ducting phenomenon received considerable attention because of its significance for submarine radars. This application never became operationally significant since submarines prefer not to radiate and subsequently did not use their radars extensively.

The next resurgence of interest in refractivity and exploitation of its effects came in the late 1960s with the advent of low flying anti-ship missiles. The early detection of such missiles beyond the normal radar horizon using radars optimized in frequency and height above the surface became a subject of intense scientific investigations and speculative radar designs. The Fleet became so concerned about refractivity effects that periodic meetings on this subject were held. Finally, in 1973, a Navy wide conference on atmospheric refractivity was convened by the First Fleet (now Third Fleet) in San Diego. The major action item was the formation of an ad-hoc group which was to formulate an organized and speedy course to provide the Fleet with a scientifically sound and operationally useful refractive effects assessment capability. The ad-hoc group convened in November 1973 and proposed as the major item the development of a shipboard, near real time propagation assessment system. The name Integrated Refractive Effects Prediction System was coined and development responsibility was assigned to the Naval Ocean Systems Center [1]. Other developments proposed at this meeting were a miniature refractionsonde for easy use on any ship and the operational implementation of an airborne microwave refractometer.

After two years of development , IREPS was tested on the USS *ENTERPRISE* (CVN 65) during a Fleet exercise. It was subsequently patented under US patent No. 4,125,893. This system provided such radically new capabilities, that it found immediate enthusiastic acceptance. Specifically, translation of refractive effects assessment into TDAs attracted entirely new user communities. The first (and still one of the most widely used) TDA is based on the IREPS radar coverage diagram shown in figure 3. This diagram shows for the specific radar selected and the measured refractiv-

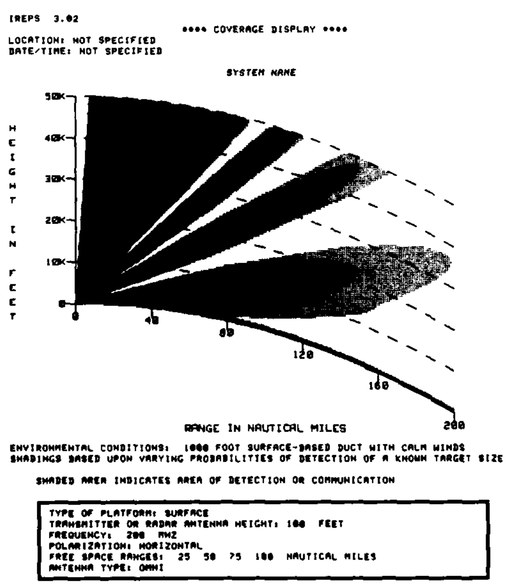


Figure 3. Radar Coverage Diagram

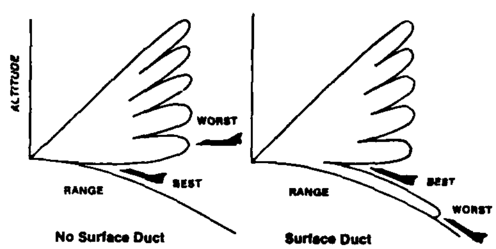


Figure 4. Tactical Use of Radar Coverage Diagram

ity conditions, detection ranges as a function of altitude and distance. In the case depicted, a surface-based duct with a thickness of 1000 feet extends detection ranges on the bottom side of the lowest lobe (the lobes in the radar detection range envelope are caused by constructive and destructive interference of the direct and the sea-reflected radar energy) and causes greatly extended surface detection ranges far beyond the normal radar horizon. The different shadings in figure 3 are for different values of detection probabilities. Figure 4 illustrates one tactical use of the radar coverage diagram. The left side of the figure shows schematically the radar detection envelope of a radar which an attack aircraft is approaching. In a non-ducting marine environment, the best flight altitude for staying undetected, is close to the ocean surface. In the presence of a surface-based duct (as in the case of figure 3), this low altitude would result in early detection and be the worst altitude to fly. The best altitude for avoiding detection is just above the duct and is readily determined from the display in figure 3.

So enthusiastic was the Fleet acceptance of this new environmental assessment capability that it was decided to proceed with the simultaneous development of both an interim shipboard capability and an operational system according to established R&D development practices. The interim IREPS, based on a desk top computer, was first installed on the USS RANGER (CV 61) in 1978, and subsequently on all deployed aircraft carriers as well as a few other selected ships.

Meanwhile, the development of the operational system named E/WEPS (Environment/Weapons Effects Prediction System) continued. It ceased as a separate development when it was decided to combine E/WEPS, underwater acoustics assessment and environmental (oceanographic and atmospheric) information in one system called TESS [6]. The first version of TESS became operational in 1986 and gradually replaced the interim IREPS. For almost 10 years, the interim IREPS fulfilled a dual role of giving the Fleet an improved war fighting capability and the R&D community a unique test bed. Under US exchange agreements, IREPS has been provided to other countries: United Kingdom, Federal Republic of Germany, Spain, Sweden, Saudi Arabia, France, Australia, Singapore, Israel, Japan; other requests are pending. Japan even funded the US Navy to develop a microwave refractometer interface and a special version of IREPS for their use.

The exchange of information on refractivity assessment between the US Navy, other US agencies and foreign countries has been essential for the successful development of the present capability. For example, without the cooperation of the National Weather Service, present data bases would not be available. The National Oceanic and Atmospheric Administration conducts promising research in radar sensing of refractivity [13]. The evaporation ducting models were developed and validated by German researchers [14,15]. The US is actively involved with various NATO research groups, and valuable information has been gained from exchange programs and joint measurement trials. These interactions illustrate the fact that propagation assessment for military applications is an interdisciplinary science requiring a broad background in atmospheric physics, mathematics, engineering and computer science, and needs the close interaction with the military user community.

Present efforts in providing an improved Navy refractivity assessment capability include development of additional TDAs, provision of refractivity data for command and control systems, millimeter wave propagation studies, improvements in evaporation ducting assessment, mesoscale meteorology modeling, development of refractivity sensing techniques including satellite techniques, development of models and techniques for simulating systems performance, and propagation modeling in horizontally inhomogeneous environments. Among the TDAs under development or recently completed are airborne radar stationing aids, airborne surface search radar coverage assessment, and a height-finder radar TDA. Since the environmental community charged with the task of providing refractivity data is not the end user of this information, another effort is underway to inject appropriate refractivity data into command and control systems. To accomplish this task connections must be established between TESS and the command and control system. Both presently-used systems such as the Flag Data Display System, the Tomahawk Weapons Control System, the Joint Operational Tactical System, and future Command Data Display Systems (CDDS) such as the Advanced Combat Direction System are addressed.

Noteworthy are recent developments in propagation measurement and modeling, evaporation ducting assessment and refractivity sensing. In millimeter wave propagation, the theoretically postulated influence of the evaporation duct on 94 GHz was experimentally demonstrated for the first time. Besides potential countermeasure applications, this frequency band may be used for missile seekers. Significant progress in evaporation ducting assessment is now possible as a result of ongoing efforts in both improvement of evaporation ducting climatologies and in situ sensing of air-sea temperature differences. A new algorithm has been developed which considerably reduces errors in existing evaporation ducting climatologies and a novel air-sea temperature sensor provides the necessary data with the required accuracy. A miniaturized radiosonde has recently become available. Promising effort is underway to develop mesoscale numerical models suitable for shipboard execution which give estimates of refractivity distributions in the vicinity of the battle group. Special emphasis is given to refractivity sensing techniques using presently available satellite sensed data. Encouraging results have been obtained in relating visible and IR imagery to atmospheric ducts. Profiling lidars can provide vertical refractivity profiles, albeit under clear sky conditions only. The scientific community has, generally, a good appreciation of the importance of environmental knowledge to naval

warfare and so does the Fleet. Where environmental considerations are often neglected is in the development and acquisition process, sometimes with costly consequences or seriously reduced performance capabilities. One key reason for this neglect is the non-availability of a simulation capability tailored toward the development community. In recognition of this shortcoming, an Engineer's Refractive Effects Prediction System (EREPS) is being developed which is specifically designed for simulating systems performance [4]. An essential part of this system is the inclusion of world-wide refractivity data bases to allow statistical performance simulation.

IREPS and TESS assume horizontal layering of refractivity for the region of interest. While extensive studies and many years of operational refractivity assessment experience have shown that this is a good first order approximation, it is obviously not correct in the vicinity of fronts, in coastal areas, and in regions of rapidly changing sea surface temperatures. A three pronged effort is, therefore, being conducted to include horizontal inhomogeneity into the Navy's assessment capability. The propagation modeling effort is reasonably well in hand. Ray tracing techniques have been developed for laterally varying refractivity conditions [16]. Both rigorous waveguide techniques and the so-called parabolic equation approximation allow accurate modeling of propagation in horizontally inhomogeneous guides [17-23]. The latter approach appears suitable for shipboard computers and will be implemented in the near future.

The critical problem for refractivity assessment under horizontally varying conditions is availability of timely and accurate three dimensional refractivity information. Future efforts will, therefore, emphasize two topics: use of satellite sensing techniques to describe the three dimensional refractivity field and improvement of numerical mesoscale models that are adequate for this purpose [24,25]. Since entirely rigorous solutions are unlikely to be available soon, empirical data have to be used as well as expert systems and artificial intelligence techniques. In addition, improved direct and remote ground-based refractivity sensing techniques need to be developed. Radiosondes and microwave refractometers will remain the major sources for refractivity profiles. Profiling lidars may be supplements to those techniques under clear sky conditions and their practicability for shipboard use will be further investigated. There is, presently, little hope that radiometric methods can provide profiles with sufficient vertical resolution to be useful for propagation assessment. There is, however, some hope that radars themselves can provide refractivity profiles through a recently derived relationship between refractivity gradients and the ratio of the variances of refractivity fluctuations over vertical wind velocity fluctuations [13]. Further TDAs will be developed in close cooperation with the Fleet and maximum utilization of refractivity data by CDDSs pursued. The EREPS simulation capability is expected to become a widely accepted tool by the engineering and acquisition community in new systems development. Propagation models will be further refined and provided to the Ocean and Atmospheric Master Library maintained by the Naval Oceanographic Office for the Commander, Naval Oceanography Command.

ELECTRO-OPTICAL PROPAGATION ASSESSMENT AND TDAS.

In the U.S., a coordinated effort between the three military services was started in 1976 to provide an assessment capability for performance of electro-optical systems. The Air Force was given the primary responsibility for development and update of atmospheric transmission codes such as LOWTRAN (low resolution transmission). The Navy concentrated on open ocean conditions with the responsibility to provide marine aerosol models [5]. The Army's primary task was the assessment of battlefield dust and smoke obscuration effects.

Extensive measurement and modeling programs resulted in a first marine aerosol model called Navy Aerosol Model (NAM) which was incorporated into LOWTRAN 6 [26]. While NAM was a first important step in the development of marine atmospheric propagation effects on visible and infrared systems, its lack of considering vertical atmospheric structure limits its applicability. A joint effort between a number of research groups under the leadership of Gathman at the Navy Research Laboratory attempts to alleviate this shortcoming through the development of a Navy Oceanic Vertical Aerosol Model (NOVAM) [27]. How important vertical structure is to assessing systems performance may be illustrated by vertical extinction coefficient profiles calculated from a NOVAM generated aerosol profile. Figure 5 shows such profiles for three different wavelengths (1.06, 3.5, 10.6 microns) calculated from data obtained during an experiment called MILDEX (Mixed Layer Dynamics Experiment) [28]. The extinction values for the three wavelengths are comparable directly at the surface but have considerably different values throughout the marine boundary layer which is capped by a temperature inversion. At the inversion height, extinction values change by orders of magnitude. This illustrates the importance of considering vertical structure if an assessment of slant path extinction is needed as in the case of an incoming aircraft or missile. Attempts to develop remote sensors (lidars) for slant path extinction have failed so far to produce devices that would be useful under operational conditions.

Tactical Decision Aids for EO systems have been developed under a program called PREOS. Figure 6 shows a display for an airborne FLIR providing detection ranges for different surface targets. In the upper portion of the figure, a table provides detection or identification ranges for various targets as a function of altitude. The same information is displayed in graphical form below the table. Comparisons between observed and calculated (derived from measured meteorological data) detection ranges have shown

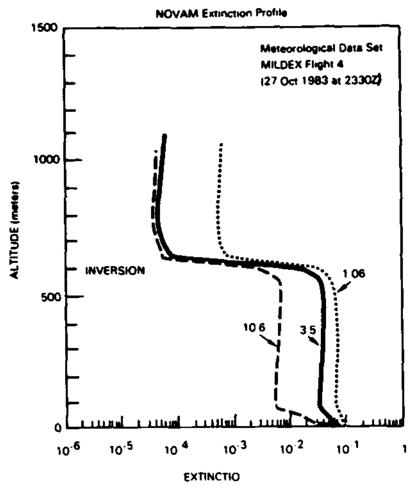


Figure 5. Extinction Coefficient Profiles for 1.06, 3.5 and 10.6 Microns Calculated from a NOVAM Aerosol Profile.

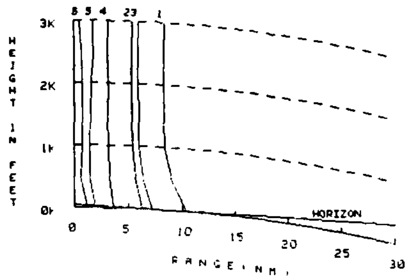
**** NOSC 02 FLIR MODEL ****

LOCATION: SN1
TIME: 2/11/81(M) -2037Z VISIBILITY= 10 NM

*** EO PERFORMANCE RANGES ***

FLIR PERFORMANCE RANGES(NM):

#	TARGET TYPES	ALTITUDE LEVELS(FT) SURFACE	500	1000	1500	2000	2500	3000
1	CRUISER/DET	10.4	9.0	8.4	8.4	8.3	8.3	8.4
2	/IDENT	6.3	5.6	5.4	5.3	5.3	5.3	5.4
3	SUBMARINE DET	7.2	6.4	6.0	6.0	5.9	6.0	6.0
4	/IDENT	3.6	3.2	3.1	3.1	3.1	3.1	3.2
5	SNORKEL/DET	2.0	1.5	1.5	1.5	1.7	1.9	1.9
6	PERISCOPE/DET	1.2	.6	.7	.8	.8	.8	.7



AIRBORNE FLIR
FAR IR BAND

PERFORMANCE RANGE CURVES INDICATE 50% PROBABILITY OF SUCCESSFUL PERFORMANCE

Figure 6. Airborne FLIR TDA

information is displayed in graphical form below the table. Comparisons between observed and calculated (derived from measured meteorological data) detection ranges have shown considerable scatter when the two ranges are plotted against each other. The reason for the scatter is our present inability to describe with sufficient accuracy the large number of contributing phenomena (illustrated in figure 2) that must be considered in performance assessment. For example, the most critical parameter for FLIR range predictions is the ship-background temperature contrast which varies with aspect, time of day, duration of deployment and atmospheric parameters. The complex task of obtaining the necessary information is described in further detail by Hughes [29]. Further improvement of TDAs for EO systems will be the result of better aerosol models, improved sensing techniques including the use of satellite sensors, and more accurate target signature codes.

IONOSPHERIC RADIO PROPAGATION ASSESSMENT AND TDAS.

The development of a real time propagation assessment system for hf and associated TDAs was prompted by a series of high altitude (118 000 km) satellites designed to measure solar radiation characteristics. In the early 1970s it was decided to develop a computer terminal which would use both data measured by these satellites (like x-ray flux, solar wind, magnetic data etc.) and provide data useful to communications and electronic warfare. The system developed was called PROPHET and became operational in 1976 [7,8]. When the high altitude solar radiation satellites were no longer available, other data sources for input parameters were substituted (e.g. x-ray flux measured by GOES (Geostationary Operational Environmental Satellite)). PROPHET was primarily developed using simple, empirical models. It proliferated into many different versions. It is estimated that some 500 terminals have been and still are in operation. Figure 7 shows a sample of a PROPHET display for signal strength in the frequency band from 2-40 MHz as a function of time. The propagation path is between Honolulu and San Diego and the systems and solar parameters are listed on top of the figure. The dashed lines show the MUF (maximum usable frequency), FOT (frequency of optimum transmission) and LUF (lowest usable frequency). This kind of analysis and display has proven to be a very helpful decision aid for communicators and has become an important part of frequency

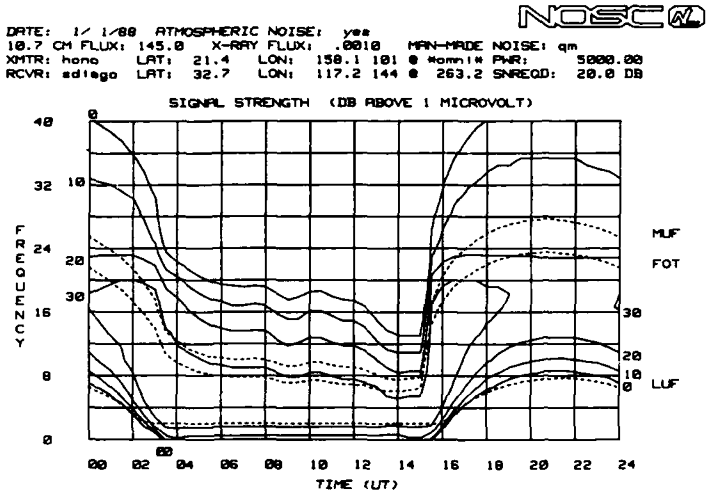


Figure 7. Diurnal Field Strength Plot for HF Propagation

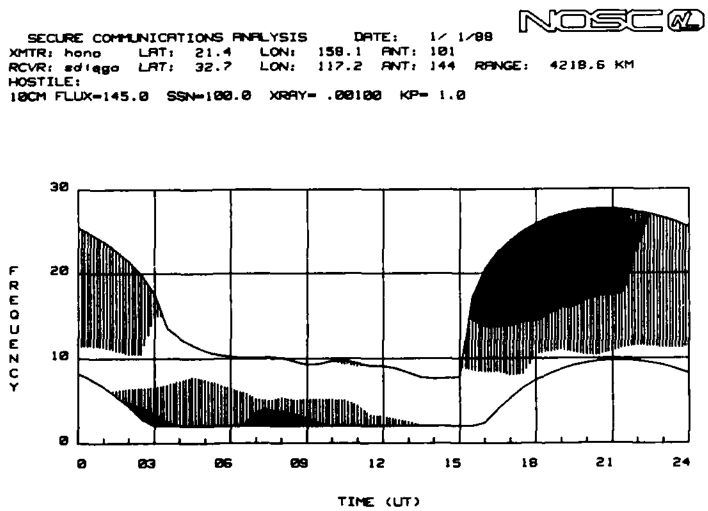


Figure 8. Secure Communications Analysis

management systems. Numerous other decision aids have been developed and successfully used. An example of a secura communications analysis is shown in figure 8. MUF and LUF are displayed for a specific propagation path (Honolulu to San Diego). The operator can choose from his library a number of hostile intercept stations. The solidly filled areas indicate frequency-time regions for which a signal cannot be intercepted by any of the previously specified intercept stations. The partially filled, lighter shaded areas indicate the possibility of a hostile intercept but no directional fix. Finally, the white areas between the MUF and LUF boundaries indicate frequency-time regions for which the signals transmitted can be intercepted by a sufficient number of specified intercept stations enabling a good position fix. Based on this decision aid, frequency selections for secure communications can be determined. Further applications and implementations of hf propagation assessment systems and associated tactical decision aids are discussed by Rose [9] and Goodman [30].

Future effort for hf propagation assessment systems will concentrate on the validation and improvement of models. Much of the past work has used empirical models whose major virtue was simplicity [31]. With increased computer capability, more complex models can be executed fast enough for near-real time applications. Also, the increasing use and availability of oblique and vertical incidence sounders makes this data source an attractive additional input for assessment systems. Finally, hf systems like over-the-horizon radars and new geolocation techniques (time difference of arrival) require a much more detailed description of ionospheric fine structure and need special attention from the modeling, sensing and assessment community.

CONCLUSIONS

Propagation assessment systems throughout the electromagnetic spectrum have been successfully developed and extensively used by the operational forces. TDAs based on these assessment systems have increased the effectiveness of our armed forces. Yet, more needs to be done since developers of assessment systems sometimes had to employ overly-simplified or empirical models to deliver a fully operational system. Examination and improvement of previously implemented models is an important responsibility of the scientific community. Availability of new environmental sensing techniques and more powerful computers offers great opportunities for future propagation assessment systems and associated TDAs.

REFERENCES

[1] Hitney, H. V. and J.H. Richter, "Integrated Refractive Effects Prediction System," Nav. Eng. J. 88 (2), pp 257-262, 1976

[2] Paulus, R.A, "Propagation-based Decision Aids in the U.S. Navy," AGARD Conf. Proc., CP 453, Paper No. 5, May 1989

[3] Barrios, A.E., "An Altitude-Error Display for Height-Finder Radar," NOSC TR 1152, January 1987

[4] Hitney, H.V., "Engineer's Refractive Effects Prediction System," AGARD Conf. Proc., CP 453 ,Paper No. 6, May 1989

[5] Richter, J.H. and H.G. Hughes, "Electro-Optical Atmospheric Transmission Effort in the Marine Environment," NOSC TR 696, May 1981

[6] Jensen, J.J., "The U.S. Navy Tactical Environmental Support System (TESS(3))," AGARD Conf. Proc., CP 453, Paper No. 4, May 1989

[7] Richter, J.H., I.J. Rothmuller and R.B. Rose, "PROPHET: Real Time Propagation Fore casting Terminal," Proc. 7th Technical Exchange Conference, Published by ASL, WSMR NM 88002, pp. 77-81, April 1977

[8] Rose, R.B., "PROPHET - An Emerging HF Prediction Technology," in Effect of the Ionosphere on Radiowave Systems (J. Goodman ed.), pp. 534-542, April 1981

[9] Rose, R.B., "PROPHET and Future Signal Warfare Decision Aids," AGARD Conf. Proc., CP 453, Paper No. 8, May 1989

[10] Kerr, D.E. (ed.), "Propagation of Short Waves," MIT Radiation Laboratory Series No. 13, 1951

[11] "Symposium on Tropospheric Wave Propagation," Naval Electronics Laboratory, San Diego, Report 173, July 1949

[12] "Report on the Factual Data from the Canterbury Project," Vols. I,II,III. Welling ton, New Zealand, Dept. Sci. Industrial Res., 1951

[13] Gossard, E.E. and N. Sengupta, "Measuring Gradients of Meteorological Properties in Elevated Layers with a Surface-Based Doppler Radar," Radio Science, 23(4),pp. 625-639, 1988

[14] Jeske, H., "The State of Radar Range Prediction Over the Sea," AGARD Conf. Proc., 70 (2), pp. 50.1-50.10, 1971

[15] Vieth, R., "Propagation Predictions for the North Sea Environment," AGARD Conf. Proc., CP 453, Paper No. 23, May 1989

[16] Patterson, W.L., "A Raytrace Method for a Laterally Heterogeneous Environment," NOSC TR 1180, Jul 1987

[17] Ishihara, I. and L. B. Felsen, "Hybrid PE-Ray-Mode Formulation of High Frequency Propagation in a Bilinear Tropospheric Surface Duct," AGARD Conf. Proc., CP 453, Paper No. 17, May 1989

[18] Dockery, G. D. and E. R. Thews, "A Parabolic Equation Approach for Predicting Tropospheric Propagation Effects in Operational Environments," AGARD Conf. Proc., CP 453, Paper No. 18, May 1989

[19] Ryan, F. J., "A Parabolic Equation Radio Propagation Assessment Model," AGARD Conf. Proc., CP 453, Paper No. 19, May 1989

[20] Craig, K. H. and M. F. Levy, "A Forecasting System Using the Parabolic Equation: Application to Surface-to-Air Propagation in the Presence of Elevated Layers," AGARD Conf. Proc., CP 453, Paper No. 20, May 1989

[21] Fournier, M., "Etude de la propagation dens une atmosphère inhomogène dans les directions horizontale et verticale par la méthode de l'équation parabolique," AGARD Conf. Proc., CP 453, Paper No. 21, May 1989

[22] Pappert, R. A., "Propagation Modeling for Some Horizontally Varying Ducts," AGARD Conf. Proc., CP 453, Paper No. 22, May 1989

[23] Levy, M. F., "Assessment of Anomalous Propagation Predictions Using Minisonde Refractivity Data and the Parabolic Equation Method," AGARD Conf. Proc., CP 453, Paper No. 25, May 1989

[24] Rosenthal, J. and R. Helvey, "Weather Satellite and Computer Modeling Approaches to Assessing Propagation Over Marine Environments," AGARD Conf. Proc., CP 453, Paper No. 47, May 1989

[25] Tag, P. M., "A One-Dimensional, Shipboard Model for Forecasting Refractive Effects in the Planetary Boundary Layer," AGARD Conf. Proc., CP 453, Paper No. 46, May 1989

[26] Gathman, S. G., "Optical Properties of the Marine Aerosol as predicted by the Navy Aerosol Model," Opt. Eng., 22, 57-62, 1983

[27] Deleeuw, G., K. L. Davidson and S. G. Gathman, "Physical Models for Aerosol in the Marine Mixed-Layer," AGARD Conf. Proc., CP 453, Paper No. 40, May 1989

[28] Gathman, S. G., Private Communications

[29] Hughes, H. G., "Airborne FLIR Detection of Surface Targets," AGARD Conf. Proc., CP 453, Paper No. 16, May 1989

[30] Goodman, J. M., "Decision Aid Design Factors in Connection with HF Communications and Emitter Location Disciplines," AGARD Conf. Proc., CP 453, Paper No. 9, May 1989

[31] Sailors, D. B., R. A. Sprague and W. H. Rix,"MINIMUF-85: An Improved HF MUF Prediction Algorithm," NOSC TR 1121, July 1986

DISCUSSION

C. GOUTELARD, FR
English Translation

I found your presentation very interesting, it shows the range of problems that will be addressed during this symposium. Problems related to propagation are of course important, but those related to jamming are important too. You said a lot about the first - can you mention the studies you are developing on the second, particularly those dealing with forecasting?

AUTHOR'S REPLY

In the tropospheric assessment systems we have TDAs which are designed to enhance or counteract jamming. Sea clutter effects, specifically when enhanced under ducting conditions, are considered in these models also. In the ionospheric assessment systems, we consider the ambient noise environment and have TDAs for signal-to-jam ratio contour displays.

HUMAN FACTORS ASPECTS OF DECISION SUPPORT SYSTEMS

by

K.F.Kraiss
Research Institute for Human Engineering (FGAN-FAT)
Neuenahrerstrasse 20
D-5307 Wachtberg-Werthhoven
Federal Republic of Germany

1. INTRODUCTION

Modern computer based systems usually confront their users with an enormous complexity. This statement applies in particular to real time or near real time applications as, for example, vehicle and process control or command,control,communication, and intelligence (C3I).While in former times there often was too little information for reasonable decision making, today in general there is too much. Hence, important detail may be overlooked in an avalanche of irrelevant detail. There is also concern with respect to the reliability of computers in such systems /1,2/. As a consequence,efforts are undertaken to support the human user of highly automated systems by providing so-called decision support systems (DSS).If we follow the definition proposed by Zachary /3/, a DSS is "any interactive system that is specifically designed to improve the decision making of its user by extending the user's cognitive decision making abilities"

In order for a DSS to be useful, users must be able to integrate the computer aid into their own cognitive processes. The success of,for example, an expert system depends, as Madni /4/ mentions, not only on the quality and completeness of the knowledge elicited from experts, but also on the compatibility of the recommendations and decisions with the end-user's conceptualization of the task. In building the knowledge base, there may also be conceptual conflicts between the knowledge engineer (KE) and the subject matter expert (SME). Obviously, there is ample opportunity for mismatches between the developer's and user's conceptualization of the domain, especially with respect to individual and interindividual differences among users. A prerequisite for DSS design therefore is the identification of what is limiting a person's decision making performance. A purely technology driven development of new automation capabilities can produce unintended and unforeseen negative consequences and should be avoided /5/.

This paper addresses human factors aspects of DSS design. Chapter 2 identifies various relevant dimensions in decision making and problems solving, followed by a discussion of characteristics and constraints in human information processing. On this basis, design goals and guidelines are identified (Chapter 3). The implementation of DSS, which is addressed in chapter 4, concerns the layout of the human computer interface, the degree of automation, as well as the selection of suitable decision aiding algorithms. It is shown that a novel systems architecture is needed to ensure cooperative task performance of the man computer team. Finally chapter 5 addresses various problems of interacting with DSS and a compilation of available operational experience. The paper ends with concluding remarks and a list of references.

2. DIMENSIONS OF DECISION MAKING AND PROBLEM SOLVING

This section addresses the structure and nature of those cognitive activities, which commonly are termed decision making and problem solving. It includes a classification of decision making with respect to various dimensions and puts forward the problem space concept with respect to problem solving.

Decision making may be defined as the choice of a particular action in a situation where various actions are possible. A typical decision situation is described by the matrix in Fig.1. The parameters in this matrix are: states of the world (e_n), available options (a_m), and consequences ($k_{m,n}$). Whenever a particular decision alternative is selected, given a state of the world, the associated consequence will follow.

Each decision consequence $k_{i,j}$ may be associated with an expected utility (EU) for that outcome. Mathematically optimal, i.e., rational decision making then means selecting the option that promises to maximize the resulting utility:

$$EU(a_i) > EU(a_k) \Rightarrow a_i \succ a_k. \qquad (1)$$

In case of risky decisions, where only the probability of environmental conditions is known, the calculation of the expected utility takes the following form:

$$EU(a_i) = \sum_{j=1}^{n} w_{i,j} * p(e_j) \qquad (2)$$

Deciding according to this maximal utility strategy guarantees optimality of course only in a statistical sense, and not for each individual decision.

From this short description, several steps of the decision making process can be identified. In the beginning information must be collected in order to structure the decision matrix. Relevant features of the state of the world like situational objectives, underlying processes, and task dynamics have to be selected. Subsequently, the reliability of risky or fuzzy data must be evaluated, for example, in terms of conditional probabilities. Finally, utility assessments for various decision options must be performed, and a choice has to be made among those options.

	e_1	e_2	.....	e_n
a_1	$k_{1,1}$	$k_{1,2}$	.....	$k_{1,n}$
a_2	$k_{2,1}$	$k_{2,2}$	.....	$k_{2,n}$
	.	.		.
	.	.		.
a_m	$k_{m,1}$	$k_{m,2}$	.....	$k_{m,n}$
	$p(e_1)$	$p(e_2)$		$p(e_n)$

Fig.1. Decision matrix. Whenever a particular decision alternative (a_i) is selected, given a state of the world (e_j), the consequence will be $(k_{i,j})$.

Some of the difficulties in making decisions become obvious when regarding the decision matrix. Firstly, the degree of completeness of the matrix determines, whether the decision task is structured or unstructured. Obviously, it is difficult to come up with a rational choice if not all environmental influences or options for action are known. Secondly, the conditions associated with the state of the world can be considered. Depending on the degree to which the state of the world is known, a distinction can be made between safe ($p_{ej} = 1$), risky ($p_{ej} < 1$), and uncertain ($p_{ej} = ?$) decisions. With respect to time it is critical to know, whether the decision is to be made on-line or off-line. In off-line operations time usually is not a particularly critical factor (e.g., management information systems). During on-line operations the amount of available time may be further specified. If only seconds are available for decision making like during vehicle control or target recognition tasks, spotaneous reactions are requested, which preclude consulting with experts. For diagnosis and planning activities usually larger intervals at least in the range of minutes are acceptable, so that concious evaluation of competing options becomes feasible (process control, error management, and C^3I). Fig. 2 summarizes the essential dimensions of decision making with respect to limitations in time, data, and degree of definition.

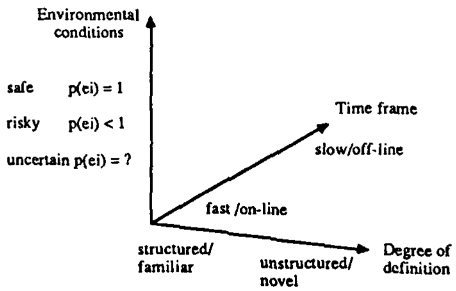


Fig.2. Essential dimensions of decision making

Problem solving is different from decision making in that it involves several steps. Starting from an actual situation, a sequence of actions (or thoughts) has to be pursued in order to reach the desired goal state (imagine, e.g., playing chess or performing a diagnosis on a car engine). The single steps involved in problem solving may be visualized as a problem space /6/. Fig. 3 shows a schematic representation.

The problem space is a graph that represents system states (nodes) and transition rules (arrows) among them. This of course must include the start and goal state. Each transition step within the problem space may take the form of an elementary decision as described above (Fig.1). Since the consequences of subsequent decisions influence each other, problem solving may be interpreted as a form of dynamic decision making. From the problem space representation it becomes obvious that problem solving requires two distinct activities. They are: problem structuring, i.e. outlining the problem space graph, and subsequently searching the problem space for a success path tcwards the goal state.

Both of these activities are formidable tasks for human beings, due to narrow limitations in memory and restricted abilities in mental imagination. The *mental problem space* representation, for example, can not be more accurate than the short term memory permits and

mental search can not reach further than the actually accessable planning horizon. The following section addresses characteristics and constraints imposed on human information processing in some more detail.

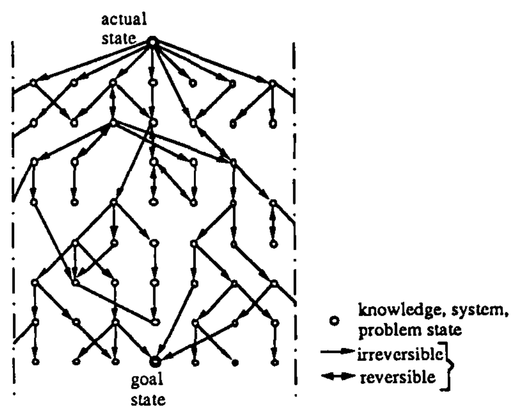


Fig. 3. Schematic representation of a problem space

3. CHARACTERISTICS AND CONSTRAINTS OF HUMAN INFORMATION PROCESSING

Following the description of rational decision making and problem solving given above, an ideal, normatively acting operator should be able to:

- establish a correct mental representation of a decision situation,
- calculate probabilities and likelyhoods associated with decision outcomes,
- quantify and scale outcome utilities consistently,
- decide rationally, i.e., select the consequence with highest utility,
- establish a mental model of the problem space,
- mentally search a problem space,
- take into account new information in an exhaustive and unbiased manner.

Obviously, such demands are not in agreement with human nature. Real operators have to operate within a framework of constraints with respect to time, knowledge, data, memory, and cognitive resources /7,8,9/. While naive decision makers at least try to act normatively, real decision makers

- act impulsively,
- use a small set of rules of thumb (heuristics), e.g.,
  - means end analysis (pursuit of partial goals),
  - representativeness (similar situations are judged to have identical outcomes),
  - availability (similarity is determined only with respect to instances available in memory),
- apply satifying criteria instead of utility maximizing criteria,
- neglect statistics and probabilities,
- have difficulty in combining competing attributes or objectives,
- use inconsistent preferences and risk assessment,
- don't consider all consequences of outcome options,
- are overly subject to situational context,
- apply bias and noise to heuristic judgement,
- have problems in analyzing or reasoning,
- have inappropriate confidence in their own decisions,
- use their internal (mental) representations whether right or wrong,
- are unable to predict processes correctly.

Experts, in general, use global not analytical assessment. They make a quick assessment of a decision situation pattern, followed by an immediate categorization of the situation and an associative decision. However, in spite of the interindividual variability in humans, a small set of recurring characteristics in decision making and problem solving can be identified. The most relevant of these are:

Goal-orientation: A decision maker unconsciously begins the decison making or problem solving process with considering the desired output of his actions. This pursuit of a goal guides information collection and hypothesis formation. It leads to expectancy, biases and cognitive tunnel vision. In addition, meaning of information is different to each person because of different training, experience, and backgrounds. What a person perceives depends,

for example, on what a person already knows. As an example, try to identify the contents of the fragmentary pictures depicted below, a task, which is related to many real world sensor data interpretation problems.

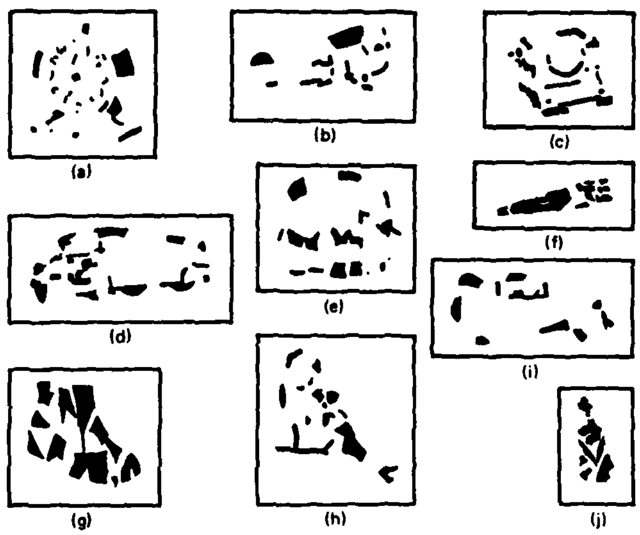


Fig.4. The effect of a priori knowledge in visual perception (Fragmentary pictures from /10/, p.205)

The reader will agree that the above pictures are hardly interpretable without context information. If additional information, concerning the possible options is given, however, (in this example the options are watch, airplane, typewriter, bus, elefant, saw, shoe, boy with dog, old cabriolet, and violine), the same task becomes pretty simple.

Mental representations: Decisions are not made using the actually available data, instead data are transformed until they fit into the available mental model. Unfortunately, mental models are mostly inaccurate, especially they are less accurate than poeple think they are. They are applied independently of the state of training or expertise an operator might have. Consequently, projections in time and space based on such mental imagery ("mind's eye") are often wrong.

Memory limitations: Reasoning can only be performed with the content of the working memory (5-7 chunks). Chunking of information is a suitable means to make the best use of this limited capacity. If additional information is required, it must be assimilated via sensors or retrieved from long term memory. Here the single channel behaviour of operator implies sequential work procedures. Also the speed of cognitive operations is limited (A mental comparison takes, e.g., typically 0.1 seconds).

Mental arithmetic: The brain is not well suited to perform complicated numerical operations. Hence, it is almost impossible to process conditioned probabilities and utilities or to estimate risks reliably in any given situation. Dealing with uncertainty by probability assessment does, for example, not follow the Bayes-theorem. Instead global estimates are used, which usually have a heavy bias.

Levels of control of human action: Depending on the task difficulty, the given time frame, and the degree of training operators show three distinctly different levels of action, namely skill based behaviour, rule based behaviour, or knowledge based behaviour /11/.In this classification skill based behaviour describes those actions, which are performed in a subconcious, quasiautomated, reactive manner. Examples are walking, playing an instrument, or ten finger type writing, i.e., activities which require instantaneous reactions. A prerequisite for skill based behaviour is long term training. Rule based behaviour can be observed in cases where the solution to a problem is known, but the single steps toward that solution must be performed consciously according to given rules. This requires no particular mental effort, but takes longer than skill based behaviour. Knowledge based behaviour, finally, is applied in situations, where no known solution to a problem exists. Consequently, creative or innovative behaviour is needed, which may take an unpredictable amount of time, while a success can not be guaranteed.

Two essential points should made here. Firstly, it must be emphasized that a particular task can be performed by a particular operator on varying levels of control, depending on his actual state of training. Also a skill achieved in a task by extensive training may be lost again if training ceases. A fallback from skilled to rule based or even knowledge based behaviour can then be observed, associated with longer operating times. Secondly, it is obvious that tasks, requiring very fast reactions, can only be mastered on the skill based level.

Human reliability: Humans are notorious for being unreliable. At a first glance the reasons for human error appear to be unpredictable. This, however, is not true, as Norman /12/ found out. He distiguishes between two types of action errors with respect to their reason, i.e., mistakes and action slips. Mistakes are made due to lack of knowledge, stress, fatigue, cognitive fixation, or wrong mental representation. In contrast to mistakes, action slips are errors made in carrying out an otherwise correct intention.

Slips occur if an inappropriate schema is activated or triggered, where schema means a sequence of linked behaviours. Several types of slips may be distinguished. *Mode errors* may occur if a situation is misclassified like when an autopilot is in a different mode of operation than expected. *Capture errors* are probable if a similar, more frequent or better learned sequence captures control, and there is a chance for *description errors* if the triggering information for a schema is ambiguous or undetected. This results in a correct operation being applied on the wrong item.

From this description it is obvious and essential to note, that slips occur mostly during periods of well trained, skill based behaviour. Since the reasons for slips are well known, slips are prime candidates for automated error prevention and compensation.

As stated in the introduction, the identification of what is limiting a person's decision making performance is a prerequisite for DSS design. Consequently, a set of design rules for DSS is listed below, which has been derived from the described human information processing deficiencies:

Improve user interface:

- match dialog form to user training and cognitive state,
- provide user guidance to reduce available options,
- increase bandwidth of data exchange by novel interaction techniques.

Support of information management:

- reduce and facilitate routine data handling work,
- provide bookkeeping functions,
- use task and situation adaptive information filtering,
- reduce memory load by visualization,
- make use of human Gestalt and figure perception capabilities.

Support decision making:

- provide decision situation structuring aids,
- perform statistics, probability and utility calculations on computers.

Support problem solving and planning activities:

- provide problem space visualization,
- provide automated reasoning functions,
- enable man-machine goal sharing.

Reduce educational and training requirements:

- provide predictor information for test and exploration,
- provide embedded training,
- provide on-line help functions.

Build error tolerance and error insensitivity into the system:

- use immediate error feedback (predictors),
- use reversibel functions (UNDO),
- use error tolerant functions (DWIM),
- make system state always unambiguously clear,
- use dissimilar command sequences for different actions,
- assign difficult actions if action consequences are severe.

The following section will address the design and implementation of DSS's. Starting from a basic DSS configuration, possible levels of automation for manual and machine functions are discussed. Requirements for a knowledge based dialog between the human and the computer are presented, and an architecture for cooperative man machine systems is developed. Finally, various aspects of information management, decision support, and error management are dealt with.

4. DESIGN AND IMPLEMENTATION OF DECISION SUPPORT SYSTEMS

Basic DSS configuration considerations:

Decision support systems usually are configured as described in Fig. 5: An operator has direct access to the system to be controlled, and in addition may exchange information with a decision support system via the user interface. As may also be seen in this figure, the DSS itself has also access to the same information about the controlled vehicle or process as the user.

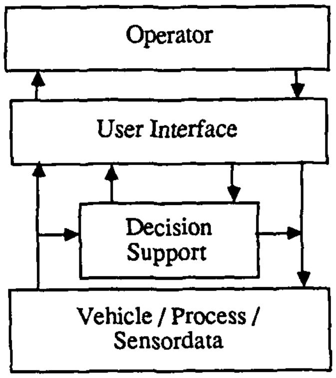


Fig.5. Basic decision support system configuration

In designing this kind of system, the following questions have to be addressed from a human factors standpoint:

- how should the interaction and authority between human and machine intelligence be organized ?
- which aspects of the users cognitive activities can be supported and how ?
- how should the user interface be designed ?

While all these aspects are essential, the question of authority assignment appears to be most critical and will be addressed first. As Seifert & Neujahr /13/ point out, the spectrum of interaction between the human and the computer reaches from full manual operation to full automation. They identify six automation levels for man-machine interface functions, which are widely accepted in the aircraft idustry, and which vary with respect to the share of authority assigned to the human (table 1):

The full *manual* and the full *automatic* modes of operation need little explanation. Some essential human factors arguments must be mentioned however: Manual operation often leads to operator stress, high workload, and fatigue. On the other hand, too much automation is likely to lead to boredom, complacency, and erosion of competence. It may be dangerous in critical situations, to move the operator too far out of the control loop by supplying too much automation.

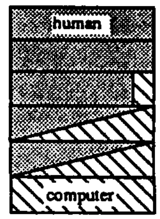
Some additional remarks have to be made with regard to the *intermediate modes of automation*, which are candidates for decision support. In the *manual augmented mode*, manual control functions are augmented by automatic control systems as, for example, nose wheel steering, and stabilizers in aircrafts. Mental decisions in this mode may be supported by electronic ckecklists, spreadsheets, integrated displays, or reversible functions for error correction. The authority remains in all these cases entirely with the operator.

Examples for the *manual augmented-automatically limited mode* are data entry formatting and validation checks as well as autonomous input error correction (do what I mean, DWIM). The corresponding approach in the area of manual control is the limiting of control inputs in order to ensure system safety (envelopes, safeguarding functioning points). Examples are angle of attack or g-level control in aircrafts, and traction control systems in cars. As depicted in table 1, part of the authority is partly transfered to the computer in this mode. The actual size of this share needs however not be fixed, but can be adapted to varying situations, system states, and user characteristics.

In the *automatic-manually limited mode* manual override and takeover of automatic functions is made possible. This is, e.g., the case with heading and course being controlled by autopilots in an aircraft. These systems are generally configured in such a way that the degree of human authority is at the decretion of the user and may reach from full automatic to full manual as depicted in table 1.

Table 1 : Automation levels for man-machine interface functions:

mode of operation	authority
manual	
manual augmented	
manual augmented-automatically limited	
automatic-manually limited	
automatic-manual sanction	
automatic	



The *automatic-manual sanction mode* is characterised by automatic functions with manual accept/reject capabilities. The assessment of options presented by an expert system, the approval of automatic target identification or priorization, and the acknowledgement of a trimming computers proposal in a submarine are typical examples. During image analysis feature identification may be performed manually (taking advantage of human vision), while knowledge based feature interpretation is delegated to the computer and reviewed afterwards /14/. The degree of human authority depends on whether manual accept/reject is optional or mandatory, and hence may cover again a wide spectrum. Sometimes manual sanction, while technically possible, is not practical due to time restrictions, as is often the case in target identification or priorization tasks.

In order to identify the relative merits of the degrees of automation outlined above, let us, in an experiment of thought, assume that the decision support be provided by a human advisor. In what way would we like his advice being presented to us ? This line of thinking leads us to look in some more detail on the ways of information exchange among humans. Nickerson /15/ has compiled a comprehensive list of conversational attributes (see Table 2).

Table 2 : Some characteristics of human conversation

Shared knowledge:	Situational context Common world knowledge Special knowledge History
Accepted conventions:	Bidirectionality Mixed initiative Apparentness of who is in control Rules for transfer of control Sense for presence Structure Characteristic time scale Intolerance for silence
Other attributes:	Nonverbal communication components Wide bandwidth Informal language Peer status of participants

From this table the crucial role of shared knowledge and accepted conventions becomes obvious. Based on this, a priori information humans are able to recognize distorted information by using shared knowledge and context. It also permits a sender to articulate ambiguously, erroneously, incompletely, and still be understood. Another attribute worth mentioning is the enormous bandwidth involved in interpersonal communication. The observation that peer status is required in order to ensure a serious conversation is of particular relevance in the context of DSS. Computer based advice will only be accepted, if it is perceived as being of exellent quality.

Certainly, some characteristics of conversation among humans would also be desirable for human computer interaction. The advice provided by a DSS should, for example, be made adaptive in the sense that it is tailored to the actual situation, to the state of the system, and to the operator behaviour and capabilities (This includes in particular sensory, motor, and mental workload as well as skill level and motivation). An intelligent assistant of this kind would allow an operator to use a system in a manner he wishes, but surrounded by a multidimensional warning and alerting system, an *electronic cocoon* /16/ that informs and supports him if he is approaching some limit (management by exception).

In order to enable a computer to behave as a humanlike partner, it must be provided additional information. Knowledge bases and procedural rules must be stored, in addition the machine must know about strategic goals and intentions. This allows the computer to check operator inputs and system outputs, and determine if they are logically consistent with the overall goal. It also provides mutual monitoring of man and machine. A system that embodies an expectancy of the *kinds of errors* that might occur may also be able to detect errors automatically as deviations from normative procedures. Systems with such capabilities are said to be *goal sharing* and *intent driven* /17,18,19/.

Cooperative human-computer system concept:

The familiar supervisory control paradigm described by Sheridan and Hennessy (1984) already contains special computer functions for operator support. The essential point is that a distinction is made between human and task interactive computers. However, in order to achieve the functionality stated above, the functionality of the human interactive computer must be improved. A possible structure for cooperative human computer systems is presented in Fig. 6. It is based on proposals made by Rouse and colleagues /21/ and Kraiss /22/. As may be seen from this figure, a set of three data bases has been introduced into the system as a complement to the human interactive computer. Stored in theses data bases is knowledge about the state of the world, about the system state, and about the human operator's state and goals.

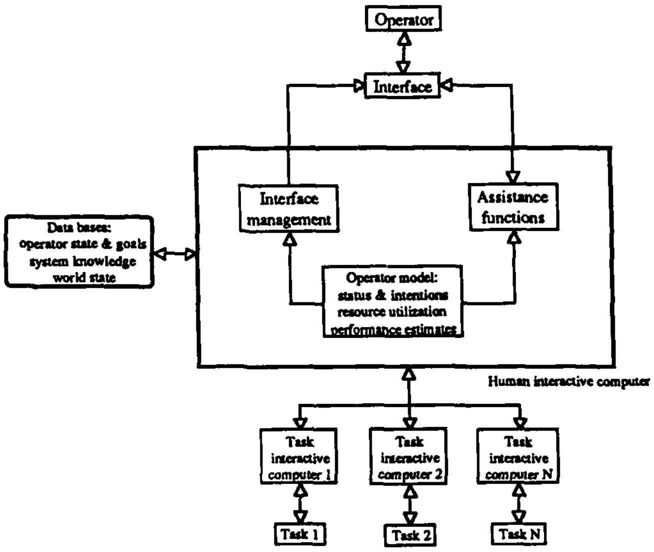


Fig.6. Cooperative human-computer system architecture

Firstly, in order to identify operator state and goals, a collection of active goals, plans, and scripts is needed. Secondly knowledge about system functioning must be encoded in specific scripts and procedures. Finally, environmental conditions and operational phases describe the state of the world. By making use of this knowledge the human interactive computer may produce intelligent and adaptive behaviour at the user interface, where adaptivity may be desirable with respect to environmental conditions, operational phases, system states, or user performance characteristics. Main function blocks of the human interactive computer are *interface management* and *assistance functions*. Both receive inputs from an *operator model*.

Interface management, i.e., the way data are exchanged at the user interface, can be shaped according to the following dimensions:

Table 3. Dimensions of interface management

information management	storage, organization, retrieval, filtering.
information display	coding, integration, selection of modality.
dialog design	task scheduling, user guidance, dialog form.

With the ever increasing amount of information available in automated systems, the role of *information management*, i.e., storage, organization, retrieval, and filtering becomes more and more critical. It is not longer acceptable to confront a user with an enormous amount of raw data. Instead, he must be supported by preprocessing and filtering functions. The approach to information filtering and alarm prioritization, taken by the Airbus-ECAM-system (Electronically centralized aircraft monitoring system), is a good example of what can be achieved in this respect (for details see /23/).

Modern computers also offer various new possibilities for *information display*, which were not available before. Among these are color and dynamics, which allow very efficient means for, e.g., the interactive evaluation of large sets of data. The amount of additional insight that is gained by colour and dynamic in interactive 3D data manipulation is impressive /24/.

Dialog design is a particularly essential aspect of interface management. Task scheduling, user guidance, and the selection of a suitable dialog form are efficient means to match the way of data exchange to the needs of the user. One especially efficient form of a dialog is direct manipulation, which allows the user to point directly to graphical objects on a screen. Fig.7 illustrates the application of this concept to the identification of ship silhouettes. As may be seen in the upper part of the figure, an observer can visually identify features and assign a frame and a name to them. In a different mode (lower part of Fig.7) the computer may initiate a dialog by graphical prompting. Both modes together make a very elegant way of interaction with pictorial data and knowledge bases /14/.

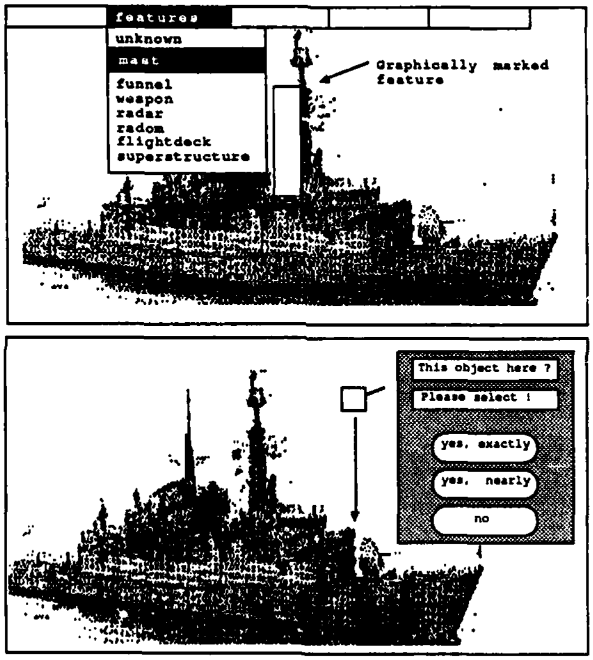


Fig.7. Direct manipulation interaction with pictorial data bases

Assistance functions may take the form of *decision support*, and/or *error management*. The spectrum of generally used algorithms in DSS is compiled in table 4. It is not the purpose of this paper to explain all these algorithms, therefore the reader is referred to the literature for details (e.g., /3,21/).

A different class of assistance functions in the human interactive computer is devoted to *error management* . The goal is to build error tolerant or error insensitive systems by the automated identification, compensation, and prevention of errors. The identification of errors is again based on the knowledge bases mentioned above. Using this information, errors of omission or commission are identified and compensated by making reference to stored legal procedures.

Table 4. Candidate algorithms for decision support

Algorithm	*Supported decision making aspect*
Process modeling	Predictions for future or different conditions.
Value modeling	Combination of attributes.
Machine learning	Compensation of systematic inconsistencies or biases for judgement refining and amplification. Adaptive aiding facilities.
Automated reasoning	Numerical techniques Calculation of conditioned probabilities and utilities. Searching for success paths in problem spaces. Symbolic techniques Generation and validation of hypotheses by rule based forward and backward inferencing.

As a simple example, consider typing errors, which often occur because not the intended key but a neighbor is activated (description error, see p. 5). Identification and correction of a typing error can, in a first step, be achieved by making reference to a thesaurus. Remaining ambiguities can be further resolved by analysing the local neighbors of an error letter on the keyboard (which are stored in the computer), and use the result for the automatic compensation of misspellings. The computer based resolution of description slips applies also to similar semantic descriptions or similarity in functionality.

If the interface management and assistance functions mentioned above are to be adaptive, input is needed with respect to the system operating phase, the state of the world, and, last not least, the operator. In particular, the identification of operator characterists is a very difficult task to be performed by an *operator model*. This refers to an operator's

- goals and intentions,
- utilization of sensory, motor, and mental resources,
- objective and subjectively perceived workload.

As Rouse and colleagues /20/ mention, the determination of an operator's goals and intentions can be achieved explicitly by asking the operator for input. Implicit determination of goals and intentions is much more difficult. One obvious source of information about the operator are - exept from physiological measures - the overt observable actions performed at the interface. The identification of intentions then requires inference processes in the context of a particular task and system situation. Reference to legal procedures which are stored in the computer also helps to discriminate expected (explained) from unexplained actions (errors or innovations).

Continuous observation is a particular interesting approach to identify individual user behaviour and preferences /25/. The basic idea is to copy user behaviour by applying machine learning algorithms (Fig. 8). As each user has a different mental model and experience, the computer can accumulate expertise from several users, making the interface smarter during operation.

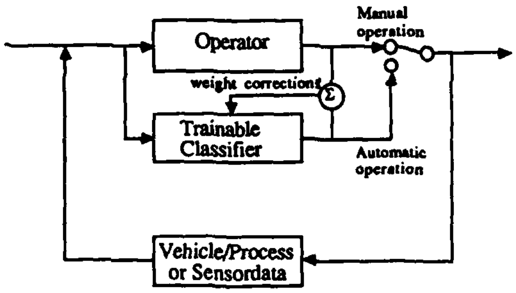


Fig. 8. Identification of operator characteristics by observation

Resource utilization (auditory/visual, verbal/spatial, cognitive, or motoric) may be derived from current and projected on-line workload analysis. Actual recource utilization may be used as a main determinant for interface management. Operator models then act as an on-line expert system for the design of displays and programmable controls. In addition, a prediction of performance in current and potential future tasks can be based on model outputs.

5. OPERATIONAL ASPECTS OF DECISION SUPPORT SYSTEMS

This section addresses various problems of interacting with DSS. Starting with an outline of user characteristics a classification scheme of DSS is presented, which takes cognitive, technological, and task related aspects into account. Subsequently, the efficieny and acceptance of decision support is discussed.

As mentioned several times in this paper, it is very essential that a DSS is matched to the cognitive processes and the mental task conceptualization of it's user. This is not a trivial task because the spectrum of operator behaviour is rather broad. Operators may, e.g., be naive or expert, either in computer handling or in the subject matter, or in both. In addition, their use of the DSS may be casual or frequent (Fig. 9). Thus, one operator may show skilled behaviour and another rule based behaviour in identical tasks.

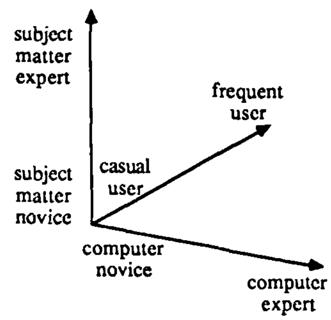


Fig. 9. Some relevant user dimensions

The actual level of action applied to a particular task depends on the task difficulty and on the available level of training and expertise (see chapter 3). Consequently, the information and support needs can not be identical for the whole spectrum of possible users. While, e.g., a casual novice will need tutoring in the subject matter as well as in computer handling, the frequent expert user needs only high level advice in the subject matter. Also a skilled operator will, after an extended break, resort for a while to rule based or knowledge based behaviour until a reasonable level of training is reached again. Table 5 shows in which ways cognitive, technological, and task related aspects of decision support systems interact.

Table 5. Cognitive, technological and task related aspects of decision support systems

Available time	Level of action	Operator task	Decision support	Properties	Implementation	
t/s 100	Knowledge based	Communicat. Navigation Diagnosis Supervision Identification	Information-mgnt Book keeping Situation structuring Visualization Risk evaluation	selfexplaining transparent adaptive cooperative	symbolic DP	sequential DP
10	Rule based	Classification Recognition Guidance	Option evaluation Teaching Consulting			
1			Display of status, command or predictor information	intuitive reactive Mode of information presentation: head-up peripheral acoustical tactile	numerical DP	parallel DP
0,1	Skill based	Stabilisation Reaction Detection	Automatic takeover			

There is also a characteristic time scale associated with different levels of action. Some tasks, like playing chess, are so demanding that, even after extensive training, they remain to be knowledge based. On the other hand there are tasks which, due to limitations in execution

time, must be performed at the skill level. Among these are reaction, stabilization, and detection tasks. Table 5 lists a number of tasks ranging from guidance to communication, which can not unambiguously be assigned to an action level. Diagnosis, e.g., can subsequently imply all action levels depending on individual strategies as well as on task characteristics.

The style of interacting with DSS is mainly determined by time considerations, and will range between the extremes *commanding* and *advice giving*. Due to the limited amount of available time during skill based action, there is no chance to begin a dialog with a DSS. The only thing that can be done is the presentation of status, command, and predictor information in a way that is intuitively correct and supportive to reactions. Alternatively, automatic limitation of control actions or automatic takeover can be considered as a means for computer aiding.

During the support of rule and knowledge based actions the time available usually allows a dialog between the user and the DSS. The support can take several forms corresponding to the levels of automation outlined in table 1. In the *manual augmented* mode the user can get mainly information management support, i.e., help in handling large amounts of information, and in structuring a decision situation. This includes task and situation dependent filtering, and display of information using advanced computer graphics features like color and animation. Other apsects of this approach is the visualization of decision and planning situations. The *manual augmented-automatically limited* mode involves no dialog with the user. Input limiting or takeover occur automatically without additional notification. The same applies to error *identification, compensation, and prevention*. Decision support in the *automatic-manually limited* mode means the delegation of tasks to automation with subsequent supervision. This is achieved simply by the selection of the desired operation modes. A real dialog between user and machine does not take place.

Decision support in the sense of advice giving, consulting, or teaching takes place in the *automatic-manual sanction* mode. The assessment of options presented by a DSS usually requires some additional inquiry about the validity, background, and context assumptions implied in the computer solution. Hence, it is not enough for a DSS to come up with a simple suggestion. Instead, it should be transparent, selfexplaining, cooperative, and adaptive to task, situations, and operational needs. Self explanation features in rule based expert systems usually take the form of long lists of rules which allow to trace back a hypothesis to causal facts or in opposite direction. However, very often this work is considered to be too tedious. Consequently, complacency and overconfidence can often be observed with operators, resulting in decision support being accepted without further check.

Therefore, it is an essential design goal for DSS to give the user as complete a picture of what is going on in the computer as possible. One obvious way to go is the substitution of textual information by graphical displays, which can be read and understood much faster. At this instance of time very few graphical DSS examples exist. One of the few is depicted in Figure 10, showing a decision parameter display for a risky decision among three alternatives. As may be seen, the computer proposes to select alternative 2, based on expected utility calculations (see equation 1).

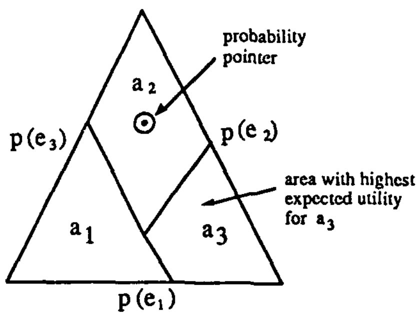


Fig. 10. Decision parameter display for a risky decision among three alternatives /26/.

However, instead of proposing "select a_2", all possible options are presented, embedded in the whole situational context. The user follows continuously the movement of the probability pointer as the environmental conditions $p(e_i)$ change, and is able to predict the instance of time, when the expected utility will become greater for a different decision. He can actually see, on which information the computer suggestion is based, and what the options and likely outcomes are.

Currently there is still little information concerning the efficiency of DSS available. The accessable data show, however, a general tendency of increased performance /27,28/:

- 2-40 % reduced response time in flight management tasks,
- 15 % faster sensor placement in intelligence task,

- improved tracking (≈10%) and target identification (≈25%) in aerial reconnaissance,
- 20 % improved consistency in pattern classification.

As far as the acceptance of DSS is concerned, reliability appears to be a very critical factor, especially if accuracy changes without warning. Eimer /29/ analysed a computer aided tracking display and observed, in agreement with the results stated above, that subjects (Ss) performed better with a high validity DSS than without it. His additional findings however are worth mentioning: Subjects performed worse with a low validity DSS than without it, and after DSS breakdown, Ss performed worse than those who never had used one. Rouse (1988) attributes this observation to the fact that "Subjects perceive their own performance as being significantly better than it actually is and expect the aid to perform much better than themselves".

Another difficulty in interacting with DSS results from the observation that decision support system functions often are not well understood by operators. Nevertheless, or may be because of this, there is a false belief in the objectivity of computer results. As an example consider the establishment of knowledge bases used in expert systems. During the acquisition of semantic knowledge a knowledge engineer (KE) extracts expertise from a subject domain expert (SME) using verbal protocols. This process is subject to biases as result of humans using cognitive heuristics as well as to inadvertant exaggerations and errors /4/. "It is in the nature of any "fact" that it is an assertion by an individual in a context, based on background of pre-understanding" as Winograd and Flores point out /30/. Hence "facts" in the computer are generated on the belief that they correspond to facts and are very likely to carry subjective biases.

As a consequence, a program written and used with (implicit) background assumptions may not be correct in a particular unforeseen situation. A rule that seems to work well in a number of different scenarios may, nevertheless, fail in a particular combination of circumstances. The reason for a wrong DSS diagnosis therefore mostly is not a programming error, but rather wrong background assumptions. It therefore appears to be mandatory that a user is sufficiently informed about the functioning of the applied decision support algorithms including the underlying assumptions. Otherwise he will not be able to critically check DSS suggestions and take over manually in case of DSS failure.

6. CONCLUSIONS

This paper addressed human factors aspects of DSS design. The spectrum of technical means for DSS implication was reviewed, in addition some example solutions were discussed. It was shown that a purely technology driven development of new automation capabilities can produce unintended and unforeseen negative consequences. In order for a DSS to be useful as a cognitive prosthetic, users must be able to integrate the computer aid into their own cognitive processes. A prerequisite for DSS design therefore is the identification of what is limiting a person's decision making performance.

Currently, there is still little information concerning the efficiency of DSS available. The accessible data show, however, a general tendency of increased performance. As far as the acceptance of DSS is concerned, reliability appears to be a very critical factor, especially if accuracy changes without warning.

A general difficulty in interacting with DSS results from the observation that decision support system functions often are not well understood by operators. There is a false belief in the objectivity of computer results. In many cases there is also an obscured responsibility distribution between man and machine, especially when cooperating with complex multiple machine experts like, e.g., the pilots associate.

Since the reason for a wrong DSS diagnosis mostly is not a programming error but rather wrong background assumptions, the interface layout is a very essential part in DSS design. It must provide the user with sufficient information about the functioning of the applied decision support algorithms including the underlying assumptions. Of course, this information must be made easily accessable and readable by using suitable display formats. Only such transparency and self explaining features will enable a critical evaluation of DSS suggestions, and a manual takeover in case of DSS failure.

7. REFERENCES

/1/ Bellin, D. and G.Chapman (1987) Computers in Battle - Will They Work ? Harcourt Brace Jovanovich, Boston,1987.

/2/ Greeley, B.M. (1988) Pentagon Vincennes Inquiry Backs Commander's Actions. AW & ST, Aug. 29, 21-22.

/3/ Zachary, W. W. (1988) Decision Support Systems: Designing to Extend the Cognitive Limits. In: M. Helander (Ed.) : Handbook of Human Computer Interaction, Chapter 47, pp.997-1030, Elsevier.

/4/ Madni, A.M. (1988) The Role of Human Factors in Expert Systems Design and Acceptance. Human Factors Special Issue: Expert Systems 30,4,395-414

/5/ Woods, D.D. & E.M.Roth (1988) Cognitive Engineering: Human Problem Solving with Tools. Human Factors Special Issue: Expert Systems 30,4,395-414

/6/ Newell, A., Simon H.A.(1972) Human Problem Solving. Prentice Hall, Englewood Cliffs, N.J.

/7/ Hink, R.L. & D.L.Woods (1987) How Humans Process Uncertain Knowledge: An Introduction for Knowledge Engineers. The AI Magazine, Fall,41-53

/8/ Wickens, C.D. & J.M.Flach (1988) *Information Processing.* In:(Wiener E.L. and D.C. Nagel eds.) Human Factors in Aviation. Academic Press, pp.263-301.
/9/ Nagel, D.C. (1988) Human Error in Aviation Operations. In:(Wiener E.L. and D.C.Nagel eds.) Human Factors in Aviation. Academic Press, pp.263-301.
/10/ Wilkening, F. (1988) Zur Rolle des Wissens in der Wahrnehmung. In Mandl H. and H.Spada (Eds.): Wissenspsychologie.Psychologie Verlags Union, München,205
/11/ Rasmussen, J. (1983) Skills, rules, knwledge: signals, signs, and symbols and other distinctions in human performance models. IEEE Tra.ıs. Syst. Man Cybern. SMC-13 (3), 257-267.
/12/ Norman, D.A. (1981) Categorization of action slips. Psychological Review, 78, 1-15.
/13/ Seifert, R., Neujahr H.(1987) Moding Strategy for Cockpit Data Management in Modern Fighter Aircraft. AGARD-GCP/FMP Symposium, Stuttgart, September.
/14/ Kraiss, K. F. and H. Küttelwesch (1987) Graphically Interactive, Knowledge Assisted Pattern Classification, in (Gilmore J.F. Ed.) *Proc. of SPIE, The International Society for Optical Engineering,* Vol. 786, pp. 592-597. Applications of Artificial Intelligence V.
/15/ Nickerson, R.S.(1976) On conversational Interaction with Computers. In: S.Treu (Ed.) User-Oriented Design of Interactive Graphics Systems. Proc. of ACM/SIGGRAPH Workshop, Pittsburgh, Pa., 101-113
/16/ Speyer, J.J. (1988) Towards Design-Induced Error Tolerance, In: Proceedings of the Int. Conf. on Human-Machine Interaction and Artificial Intelligence in Aeronautics and Space, Toulouse-Blagnac, 28-30 September, 69-94
/17/ Wiener, E.L. (1988) Cockpit Automation. In:(Wiener E.L. and D.C. Nagel eds.) Human Factors in Aviation. Academic Press, pp.*263-301.*
/18/ Shalin, V.L., Perschbacher D.L. & P.G. Jamar (1988) Intent Recognition: An Emerging Technology, In: Proceedings of the Int. Conf. on Human-Machine Interaction and Artificial Intelligence in Aeronautics and Space, Toulouse-Blagnac, 28-30 September,125-137
/19/ Onken, R. (1988) Automatic Pilot Aid for Situation Awareness and Fligt Planning, In: Proceedings of the Int. Conf. on Human-Machine Interaction and Artificial Intelligence in Aeronautics and Space, Toulouse-Blagnac, 28-30 September,125-137
/20/ Sheridan, T.B. & Hennessy, R.T.(Eds.).(1984). Research and Modeling of Supervisory Control Behaviour, Report of a Workshop. Washington: National Academy Press.
/21/ Rouse, W.B.; Geddes N.D. & Curry, R.E. (1987). An Architecture for Intelligent Interfaces: Outline of an Approach to Supporting Operators of Complex Systems. In: G.Salvendy (Ed.), *Human Computer Interaction, Vol.3* (pp. 87-122). London: Lawrence Erlbaum .
/22/ Kraiss, K. F.(1985) Fahrzeug- und Prozeßführung, kognitives Verhalten des Menschen und Entscheidungshilfen. *Fachberichte MSR 11.* Springer Verlag.
/23/ Airbus Industries (1984) Electronically Centralized Aircraft Monitoring System. Toulouse
/24/ D^2 Software, Inc. (1988) MACSPIN Graphical Data Analysis Software, Austin, Texas.
/25/ Freedy, A., Madni A., and Samet M. (1985) Adaptive User Models: Methodology and Applications in Man-Computer Systems. In: Rouse W.B. (Ed.): *Advances in Man-Machine Systems Research, Vol. 2,* JAI Press Inc., Greenwich, Conn., 249-293.
/26/ Amey, D.M.,Feuerwerger P.H. and R. Gulick (1979) Dokumentation of Decision Aiding Software: OPINT Users Manual. Decision and Designs, I. McLean, VA.
/27/ Rouse, W.B. *(1988) Adaptive Aiding for Human/Computer Control. Human Factors Special* Issue: Expert Systems 30,4,395-414
/28/ Kraiss, K.F. (1982) Assessment of Interactive Visual Pattern Classification with Adaptive Computer Aiding. *Proceedings of the 6th Int. Conf. on Pattern Recognition, Vol.2* . New York: IEEE, 944-946.
/29/ Eimer,E.O. (1987) When Decision Aids Fail, Harry Armstrong Aerospace Medical Research Laboratory, WPAFB, Ohio 45433-6573, AAMRL-TR-87-035
/30/ Winograd, T & F. Flores (1988) Understanding Computers and Cognition, a new Foundation for Design. Ablex Publ. Comp., Norwood, New Jersey

THE TACTICAL ENVIRONMENTAL SUPPORT SYSTEM (TESS(3))

Captain Jack J. Jensen, U. S. Navy
Program Manager, Environmental Systems Program Office
Space and Naval Warfare Systems Command
Washington, D. C. 20363-5100

SUMMARY

The U. S. Navy has taken an aggressive step toward improving the analysis and prediction of the performance of tactical sensors and systems at sea. The TESS(3) is one of three key components now under development which will form the basis for state of the art on-scene tactical support. Along with a high resolution satellite receiver/recorder and the automated Shipboard Meteorological and Oceanographic Observing System (SMOOS), TESS(3) is expected to significantly improve the ability of the tactical commander to exploit potential advantages which accrue from a quantitative knowledge of the surrounding environment.

INTRODUCTION

This is a dynamic period in the history of warfare. Technological knowledge is expanding at an ever increasing pace, and new products are appearing daily, yet the U. S. Navy's ability to develop and acquire systems which employ the latest technology, the acquisition process, has not always been able to respond as quickly as the fleet would like. New technology which does find its way to sea changes the nature of interactions in surveillance and battle. That employed by the enemy changes the threat and increases the difficulty in achieving success in war.

Fleet requirements for environmental knowledge have evolved considerably from the pure "weather" analyses and forecasts of forty years ago. Whereas the aircraft and vessels of that time were sensitive to the effects of the weather, subsequent generations of machines were more powerful, more capable, and thus demanded less concern for weather on the part of the operators. In the late sixties and seventies, there were increasing efforts to model the environment and apply it to the task at hand and to quantify the effects on system performance. It is this modeling which forms the basis for today's system performance assessments, such as the Integrated Refractive Effects Prediction System (IREPS), which are routinely provided to fleet commanders.

In generating the environmental support, the U. S. Navy has evolved from the manual analyses and printed nomographs which were prevalent as recently as twenty years ago, to an increasing reliance on numerical support. The mainframe processors which analyzed global conditions were installed in our primary processing center in Monterey in the early sixties. At that time, decentralization was not practical because the costs of maintaining multiple computer centers could not be justified. Thus, primary effort was focused on making the global center in Monterey as capable as it could be. This strategy permitted all resourcesto be put to optimal use and ever larger computers were acquired to run increasingly sophisticated environmental computer programs. As the quality of the product became more obvious, more interest was generated by the fleet toward receiving more products, so the system, and demands on it, grew.

Eventually, any system supporting naval units worldwide which relies on a single shore facility will suffer from the difficulties of communicating the information to tactical units in a timely manner. Fleet commanders always want to know the latest information, and they want to determinethe limits of validity for system performance predictions. In order to relieve the heavy demand on communication links, it is clearly preferable to provide the tactical commanders with their own capability to assess the environmental impacts.

Fortunately, in the early 1980's, the state of technology had progressed to the point where significant computing power could be placed on a desk top. In this case, the unit we used was the Hewlett Packard 9845 and the initial purpose was to calculate radar propagation. Other units were in use in the acoustic world, but generally the state of the art at that time could be characterized as computer limited and manpower intensive.

THE EVOLUTION OF TESS

As the 1980's progressed and the 9845's aged, they were replaced by the first desktop system specifically procured to bring real time support to the tactical decision maker in evaluating the environment. Based on a HP-9020, and developed by the Commander, Naval Oceanography Command in Bay St. Louis, MS, the Tactical Environmental Support System (TESS 1) was a significant improvement. The power of the 9020 provided a significant increase in capability aboard ship, but environmental analyses and applications remained a manpower intensive process. TESS 2, a natural extension of the TESS 1, is also hosted on the HP 9020. It has interfaces to receive broadcast radio teletype data, and can display products on the shipboard closed circuit television (CCTV). Also, an interface to a low resolution environmental satellite receiver recorder is under development.

In 1985 the Oceanographer of the Navy consolidated many of the diverse projects which he sponsored in one office within the Space and Naval Warfare Systems Command in Washington, DC. This new Environmental Systems Program Office was charged with the responsibility to plan, develop, and acquire the fully capable TESS(3) system, along with a new, high resolution environmental satellite receiver/recorder, and Shipboard Meteorological and Oceanographic Observing System (SMOOS), and a number of other systems and associated software.

The Navy awarded a contract to develop the TESS(3), the mainstay of a new architecture for oceanographic support, in July 1988 to the Lockheed Missiles and Space Company, Inc., Austin Division, which had teamed with the Lockheed Research and Engineering Center, Huntsville, Alabama, and the MacDonald Dettwiler and Associates, Ltd. of Vancouver, British Columbia, Canada.

TESS(3)

Functionally, the TESS(3) ingests data from four sources: (1) local environmental sensors, e.g., automatically from shipboard sensors or from observations keyed in manually; (2) high resolution environmental satellite data captured directly aboard ship; (3) gridfield and other data from shore-based regional oceanography centers; and (4) facsimile and standard worldwide radio teletype data broadcasts. In addition, the TESS(3) has extensive atmospheric and oceanographic climatological data bases to which it can refer, or revert if need be, at any time. All of the raw and processed data received is then "synthesized" into an optimal "nowcast" to define the three dimensional atmospheric and oceanographic environments in the region of interest. In the case of the carrier battlegroup, for example, the meteorological analyses extend 1500 km from the carrier. TESS(3) thus provides up to date environmental information in all areas of tactical interest, including that of the outer air battle.

After the environmental conditions are determined, TESS(3) prepares the usual types of environmental analyses. Although not part of the initial system, TESS(3) will eventually run limited prediction models. Perhaps the most value occurs as TESS(3) applies the real time conditions to parameters of tactical interest and assesses the impact which the environment will have on tactical operations. For example, the atmospheric environment will be analyzed to determine the affect it will have on the propagation range and coverage of particular electromagnetic frequencies. The results can be used by the tactical commander to improve his decision making with respect to how he employs his fleet assets, and what frequencies he uses. TESS(3), however, will not be able to tell him rules of engagement, change the assets available to him, or change or assess the probabilities of events like false alarms, counter detections, etc. Those are clearly the responsibility of the tactical decision maker himself.

In addition to supporting the human decision maker, TESS(3) will also support shipboard systems automatically, either in response to real time requests or in accordance with a predetermined schedule.

TESS(3) CHARACTERISTICS

The winning architecture is based on a MASSCOMP (now Concurrent) micro supercomputer which is optimized for real-time data acquisition and processing. A powerful combination of tightly coupled multiple central processing units, it is ideal for both technical computing and the manipulation of images. The design is also rugged and is intended to satisfy the military requirements for ruggedness (temperature, shock, vibration, power, TEMPEST, EMP/EMI, etc.). The majority of the hardware fits within a single cabinet, although the remaining cabinets from earlier TESS versions will serve as chasses for the imaging terminals. The software is predominantly commercial, with modules linked together in an integrated manner to provide satellite processing; a universal data base; graphics and imaging; and communications. Operators will have high resolution monitors (1024x1024 bit with the number tailored to the specific requirement); 1.2GB of on-line storage; 16 MB of system memory; and a picture quality, high resolution black and white printer.

TESS(3) has an open architecture founded on industry standards which will permit easy resupply of components and improvements to future capability, as required, with minimum effort. The central processing units are Motorola 68030s, and the architecture includes both Multibus and VME bus. Realtime UNIX AT&T System V is the operating system. Interfaces are generic, insofar as possible with particular emphasis on the IEEE 802.5 fiber optic capability.

TESS(3) is planned to be operational in mid-1991 and to be installed on 44 ships and 28 shore sites.

While the contractor is developing the baseline system, we within the Navy are continuing to rehost additional software recently developed for TESS 2 and to develop and test new tactical applications which will be integrated into the TESS(3) system shortly after it becomes operational.

CONCLUSION

For a time, TESS(3) is expected to be the most powerful computer in the fleet when it becomes operational; yet it will also be one of the least costly and most enduring. It is, however, the extension of the TESS(3) concept into the future Navy/DoD applications that will prove most exciting. Today we are limited to processing essentially one dimensional acoustic and radar propagation models, i.e., we assume that the environments are homogeneous in time and space. With the TESS(3) data base and computational power, we can quickly move into range dependent modeling, in real time, and at sea. Eventually, three dimensional models will become reality. With TESS(3), we can more fully exploit a wide range of environmental satellites already on orbit.

TESS(3) is a key element in an exciting new environmental support architecture which will provide significant improvement to the state of the art of assessing and forecasting atmospheric and oceanic environmental conditions at sea. Of equal importance, the tactical decision maker will be able to quantify the effect of the environment on the performance of his platforms, sensors, and weapons systems, and to assess the probable limits of application. TESS(3) will thus foster development of better assessments and solutions to tactical problems, and will do so in far less time than such tasks take today.

Propagation-based Decision Aids in the U. S. Navy

Richard A. Paulus
Ocean & Atmospheric Sciences Division
Naval Ocean Systems Center
San Diego, CA 92152-5000

The U. S. Navy has been using a shipboard radar propagation assessment system for the past decade. This system was conceived as the Integrated Refractive Effects Prediction System (IREPS) in 1973; tested at sea in 1976; and installed on most capital ships beginning in 1978. IREPS provided two types of products: (1) displays of refractivity data and (2) sensor performance displays. The workhorse display was the radar coverage diagram used by the air wing to determine penetration or jamming altitudes against hostile radars. This initial capability to exploit propagation effects was received so enthusiastically and proved so successful that the development of Tactical Decision Aids (TDAs) became part of an ongoing program to enhance this capability in the fleet. These TDAs structure the propagation information for the decision maker and perform functions that would otherwise overwhelm him. The decision maker is not directed to a specific course of action but rather is provided a framework within which he can make tradeoff decisions with respect to propagation in conjunction with other essential factors of his mission. This approach to the development of an aircraft stationing aid is discussed along with an overview of several TDAs applicable to various warfare areas. Efforts to incorporate these tactical decisions aids into Navy sea-based command and control systems are explored.

1. INTRODUCTION

In 1973, the need for an on-board propagation assessment system for the U.S. Navy was expressed. The concept was for a system that would utilize on-scene meteorological data to assess the effects of actual propagation conditions on system performance and aid in decisions to exploit or mitigate these effects. In 1976, a prototype called the Integrated Refractive Effects Prediction System (IREPS) [1] was demonstrated at sea aboard the USS ENTERPRISE. The IREPS assessments were very well received and an interim assessment system based on a commercial desk-top computer was installed on U.S. Navy aircraft carriers commencing in 1978. This interim system provided both an initial capability for the fleet and an avenue for feedback from the fleet to refine existing products and influence development of new products.

The initial products in IREPS included displays of refractivity data (Fig. 1) and sensor performance (Fig. 2). Figure 1 shows plots of refractive data, evaporation duct characteristics, and a general summary of propagation conditions for three geometries. Figure 2 is a coverage display showing areas of detection at differing detection criteria (target radar cross section, probability of detection, etc.) for a naval radar. The coverage display was the most utilized product in the fleet, in part because of its similarity to the fade chart with which many naval officers were familiar. However, IREPS allowed the user to see the variation of radar coverage with environmental conditions (Fig. 3) and make tactical decisions accordingly. The users could also make changes in detection criteria quite easily. The success of the coverage display prompted requirements for specialized products for other tactical applications, including airborne early warning, electronic warfare, and height-finder radars. Tabular displays of system vulnerability and surface-search radar detection were also developed [2]. A tactical jamming decision aid was the subject of a master's thesis by a naval officer student at the Naval Postgraduate School (Fig. 4). The Electronic Countermeasures Effectiveness Display shows plots of signal strength relative to free space versus height at five ranges. These plots assist in the selection of optimum jamming altitudes under varying environmental conditions. Under normal conditions, jamming aircraft would fly high (in the optical region) and aircraft trying to avoid detection would fly low (below the normal radar horizon). In the case of figure 4, jamming at a low altitude is an option while aircraft trying to avoid detection would want to fly somewhat higher than normal (above the effects of the surface-based duct).

These IREPS products were eventually included in the Navy's Tactical Environmental Support System (TESS) as the IREPS desk-top computers became obsolescent.

2. INITIATIVE IN TACTICAL DECISION AIDS (TDAs)

Because of the strong interest in tactical applications of propagation information, an exploratory development effort was initiated in 1983 that had the development of tactical decisions aids as one of its aims. Specifically, the task is to develop algorithms and displays of relevant propagation information and effects needed for tactical decisions related to surveillance and communication systems and aircraft and ship platforms. Barnes [3] reviewed several successful military decision aids for common attributes. Following his study results, the design approach taken for propagation TDAs was to structure the decision process, perform functions that could overload the user, and include the user in the development process.

One of the first propagation decision aids developed under the tactical decision aids initiative was in response to an Atlantic Fleet meteorological requirement for CAP

(Combat Air Patrol) Station Options based on refractive effects. An airborne coverage display like figure 5 was already available, but the display was for a specified altitude. This forced the user to generate displays at different altitudes to estimate the propagation effects as a function of radar altitude. An alternative to the height versus range display is a height versus height display for a specified range. Such a display is shown in figure 6, based on ray optics calculations. This display shows the refractive distortion at a range of 400 km. The shaded area in the plot corresponds to radar/transmitter and target/receiver altitude combinations for which detection, communication, intercept, or jamming may occur, dependent upon specific system characteristics. The fully shaded square between 4700 and 5000 m indicates the radar/transmitter and target/receiver altitude combinations which are within the duct, implying multipath with possible signal fading in addition to over-the-horizon propagation at greater ranges. The crescent-shaped clear area in the middle of the display arises from the effects of the trapping layer. This clear area corresponds to altitude combinations which occur within the radar/radio hole. There is reduced detection, communication, intercept, or jamming capability for altitude combinations within the clear areas. The more heavily shaded area along the lower left of the clear area indicates multipath propagation due to the presence of refractive layers. The clear area in the extreme lower left of the display represents those altitude combinations which are below the radar/radio horizon. The bar chart on the right indicates the type and vertical extent of the non-normal gradients in the refractivity profile.

As an example of the use of the AAW (anti-air warfare) Aircraft Stationing Aid display, a radar at 6 km that could normally detect targets at a range of 400 km would have reduced detection capabilities for targets between 7 and 9 km at that range. Analogously, an aircraft at 6 km would have difficulty in communicating with aircraft in the same altitude interval. For a jamming scenario, a jammer at 6 km altitude at 400 km range could jam a radar receiver at 6 km altitude. However, if the radar were moved to an altitude of 8 km, the jamming effectiveness would be reduced.

The intended users of this display are the Naval Flight Officers of the airborne-early-warning community. These officers operate the radar of the Navy's carrier-based E2-C Hawkeye aircraft. Most of these officers were familiar with displays such as figure 5 and a few had become familiar enough with refractive effects to author tactical memoranda or technical articles [4]. One individual with extensive experience dealing with the operational impact of refractive effects and personnel from Airborne Early Warning Training Squadron One Hundred Ten were asked to review and comment on the AAW Aircraft Stationing Aid at various stages in its development. These reviews substantiated the initial concept and provided insight into the varied tactical applications of the display.

A second TDA developed under the initiative is the Altitude Error Display for radars with a height finding capability [5]. Target height is not measured directly by a radar; rather, target elevation angle and range are measured and height is calculated. This calculation of height requires some assumption about the refractive structure of the atmosphere. If the actual atmosphere in which the radar is operating differs from the assumed atmosphere, then the target position indicated by the radar will differ from the actual position of the target. This position error increases with range. The contribution of the error in range to the position error is small and operationally insignificant. The error in height is by far the dominant component of the position error and can be extremely large (> 50%) in ducting situations.

The Altitude Error Display was developed to provide radar operators an assessment of refractivity induced altitude errors. Figure 7 shows an Altitude Error Display where the shading of the detection envelope is related to the magnitude of the error. For this refractive environment, a low-flying target detected at 150 m and 320 km in range would be indicated by the radar as being at an altitude of approximately 6000 m. This error complicates the problem for the airborne intercept controller at the radar console in giving proper vectors for the aircraft under his control to intercept the target. In another example, an aircraft flying level at 6000 m directly toward the radar would appear to be descending. Conversely, an aircraft flying away from the radar would appear to be ascending. Thus, it is possible for two radars on different ships tracking the same aircraft to provide contradictory information about its flight profile.

A third TDA was developed under the initiative in response to a meteorological requirement originated by the Commander Naval Oceanography Command to support airborne surface-search radars. The propagation algorithms in IREPS, and subsequently TESS, did not account for sea clutter, expected to be important relative to small targets, nor distributed radar cross section targets like ships. Sea clutter can be estimated by an existing sea clutter model [6] modified for evaporation ducting conditions [7]. Combatant-size targets can be modeled in two ways. First, the entire radar cross section can be assumed located at a height one-third the way up the superstructure. Second, the cross section can be distributed with height and the power returned integrated over the height of the target [8]. The maximum detection ranges assessed utilizing these models under varying ducting conditions are in close agreement. However, well within the radar horizon, the distributed target has the effect of filling in the nulls in the optical interference pattern that would be observed with a point target. This is a more realistic representation of propagation, but at the expense of increased time to integrate the power returned.

Figure 8 is a display of signal-to-noise ratio versus range. The horizontal dashed line is the visibility factor. Detection is implied wherever the signal-to-noise ratio

exceeds the visibility factor. The short-dashed lines are bounds of sea clutter return (± 5 dB from the average sea clutter). Sea clutter will inhibit detection of the target wherever it exceeds the signal-to-noise ratio. In the case of figure 8, the target would be detectable in the range interval 60 to 130 km, with fades associated with the optical interference pattern. Figure 9 is a display of received power versus range for a combatant-size target with a distributed radar cross section. This target is detectable at all ranges less than approximately 150 km. Received power far exceeds the sea clutter return. Also, the multipath lobes and nulls are smoothed out by the effects of the distributed target. Another option in displaying this type of propagation information is Figure 10. Here the user can see that there is an altitude interval between 500 m and 2000 m in which detection is possible. Above this interval, sea clutter will mask the target, and below this interval, the target will be over the horizon.

3. TDAs IN COMMAND AND CONTROL SYSTEMS

Command and control is the process by which a commander exercises authority over and directs the actions of his assigned forces in the accomplishment of his mission. Various communications and information systems provide support to the commander in this process. On the battle group scale, there are several areas within existing and planned command and control systems that could benefit from propagation information. So, concurrent with the initiative in TDAs, there is an effort to incorporate environmental information (including TDAs) into command and control systems for warfare commanders in the battle group. On ships so equipped, TESS will provide this support. On other ships, systems like ACDS (Advanced Combat Direction System) and JOTS (Joint Operational Tactical System) could provide propagation displays and decision aids. JOTS, currently in the fleet, already has propagation algorithms and displays. ACDS, a developmental system, is being analyzed with regards to current plans for utilization of environmental information and future enhancements. Issues concerning what information is relevant to each warfare commander, what environmental functions should be performed by a given system, and the source and extent of environmental data are being addressed.

On the scale of a single platform, the P-3C Update IV software for the maritime patrol aircraft will include a propagation assessment and TDA display capability. Other systems that could benefit from this type of capability are the E-2C and the AEGIS Weapons System. The E-2C currently has an environmental data collection capability with the airborne microwave refractometer. On-board processing of this data into appropriate tactical decision aids would enhance the E-2's effectiveness.

4. SUMMARY

Ten years ago, the U. S. Navy saw the introduction of a near real-time propagation assessment capability into the fleet. This initial capability was gradually expanded to support specific tactical applications in anti-air warfare, anti-surface warfare, and electronic warfare. Involvement of fleet users in this expansion smoothed the way for acceptance of propagation decisions aids in the tactical communities. The means to provide similar capabilities to the warfare area commanders is currently being investigated. The future may bring a truly real-time assessment capability. The microwave refractometer is already onboard many airborne-early-warning aircraft. The maritime patrol aircraft can potentially utilize a dropsonde to gather on-scene refractive information in support of their mission. Navy surface ships could take advantage of the recently introduced Mini-Rawinsonde system which can provide the profile information for the ship to assess refractivity-induced altitude errors for height-finder radars.

5. REFERENCES

[1] H. V. Hitney and J. H. Richter, "Integrated Refractive Effects Prediction System (IREPS)," Nav. Eng. J., vol. 88, no. 2, Apr. 1976, pp. 257-262.

[2] R. A. Paulus and H. V. Hitney, Naval Ocean Systems Center, "Surface Duct Effects on Radar/ESM Detection Range," 1979, NOSC TD 260, Proc. Conf. on Atmospheric Refractive Effects Assessment, pp. 69-73.

[3] M. J. Barnes, Naval Weapons Center, "Review of Five Military Decision Aids," 1980, NWC TP 6171, pp. 5-6.

[4] K. D. Denbow, Carrier Airborne Early Warning Squadron One Hundred Sixteen, "Operational Employment of Airborne Microwave Refractometer," 1979, NOSC TD 260, Proc. Conf. on Atmospheric Refractive Effects Assessment, pp. 29-32.

[5] A. E. Barrios, Naval Ocean Systems Center, "An Altitude-Error Display for Height-Finder Radar," 1987, NOSC TR 1152.

[6] M. M. Horst, F. B. Dyer, and M. T. Tuley, "Radar sea clutter model," Proc. IEEE Int. Conf. on Antennas and Propagation, London, Nov. 1978, pp.6-10.

[7] R. A. Paulus, "Evaporation Duct Effects on Sea Clutter," IEEE Trans. Ant. & Prop., submitted for publication, 1988.

[8] K. D. Anderson, Naval Ocean Systems Center, "Surface-search Radar Performance in the Evaporation Duct: Global Predictions," 1986, NOSC TR 923 Revision A, pp. 4-10.

6. ACKNOWLEDGEMENTS

The development and transition of these tactical decision aids to the fleet was a result of the combined efforts of all personnel of the Tropospheric Branch of the Ocean and Atmospheric Sciences Division.

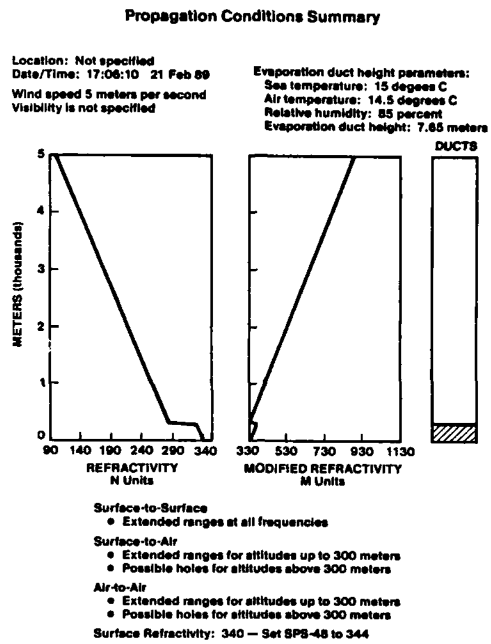


Figure 1. Example Propagation Conditions Summary from IREPS.

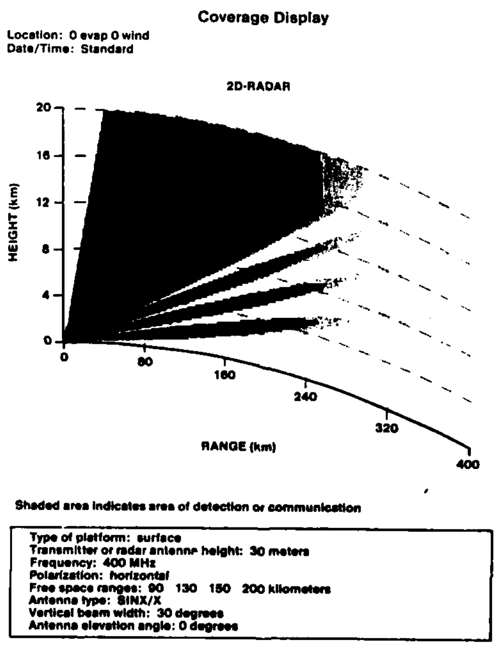


Figure 2. Example IREPS Coverage Display for standard atmosphere propagation conditions.

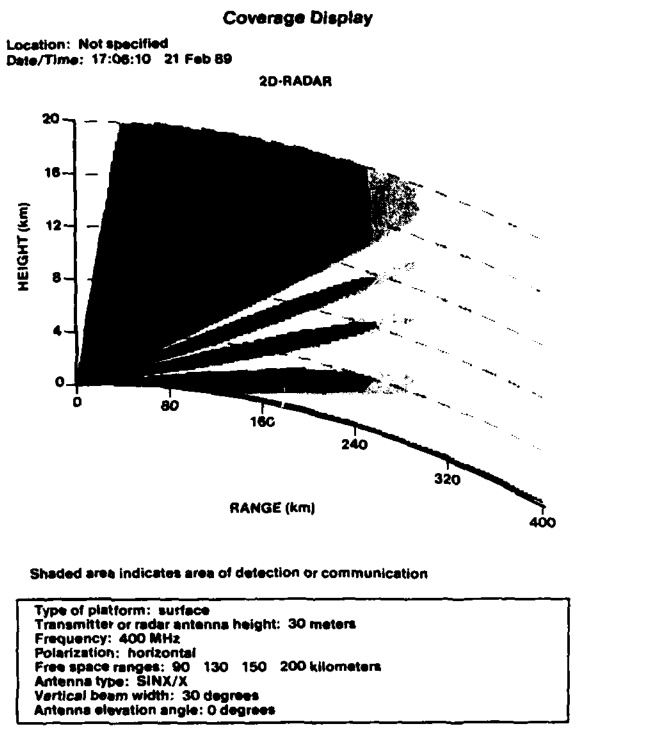


Figure 3. Example IREPS Coverage Display for a surface-based duct environment. Note the extended area of detection along the earth's surface.

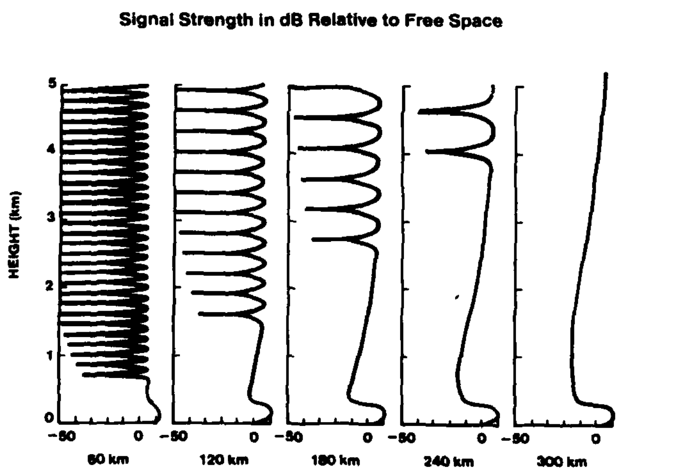


Figure 4. Example Electronic Countermeasures Effectiveness Display for surface-based duct propagation conditions.

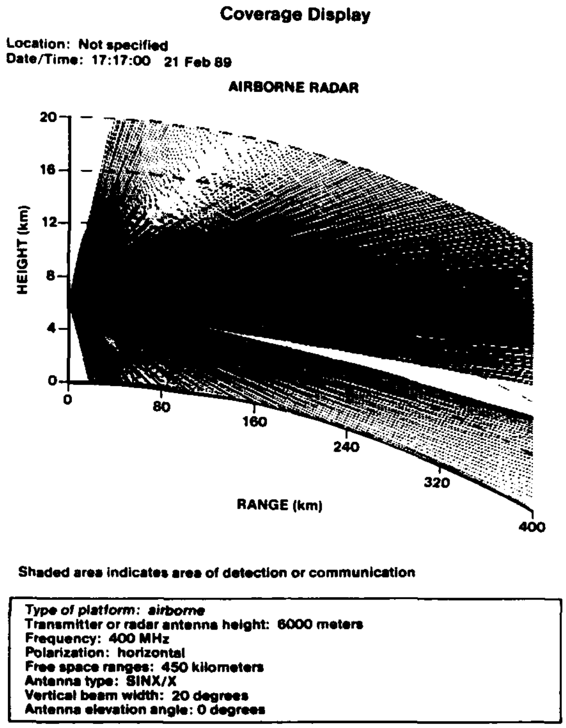


Figure 5. Example Airborne Coverage Display for a radar/transmitter at 6000 meters, above an elevated duct whose top is at 5000 meters. A "radar hole" begins at approximately 120 kilometers.

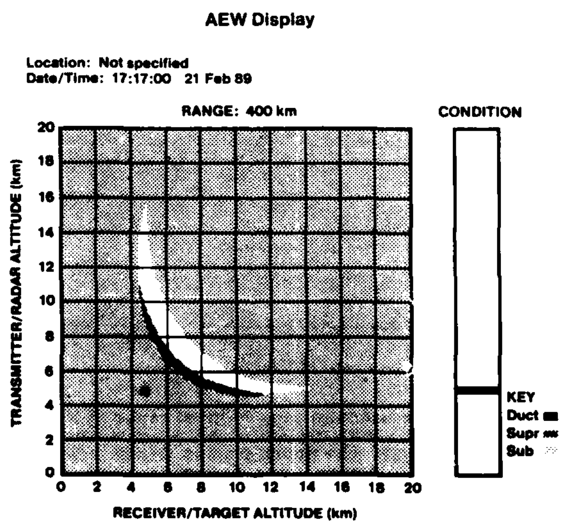


Figure 6. Example height versus height display for a range of 400 kilometers in the presence of an elevated duct.

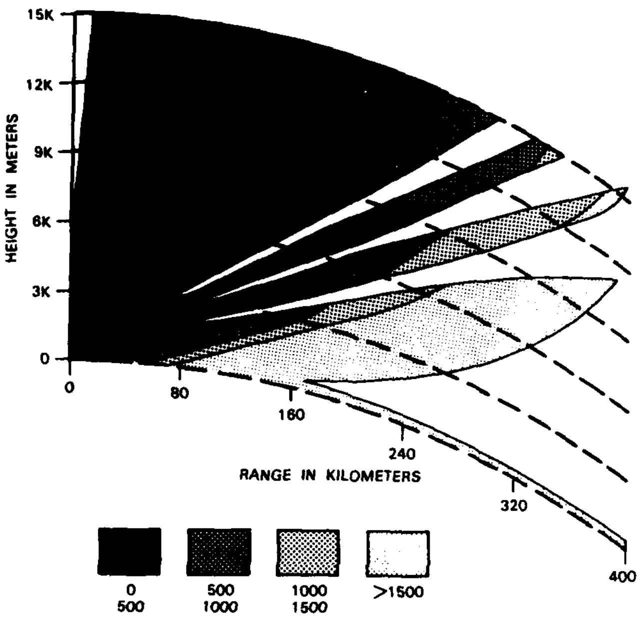


Figure 7. Example Altitude Error Display for a surface-based duct environment. The shading indicates the magnitude of the altitude error.

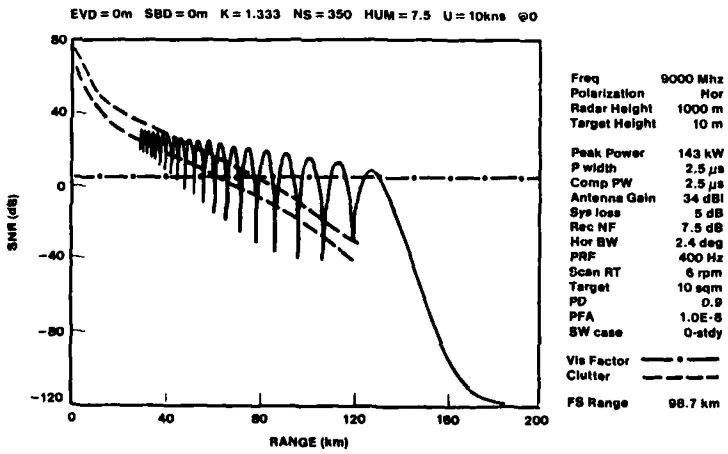


Figure 8. Example signal/noise versus range display for airborne surface-search radar detection of a small surface target. The solid line is signal, the short-dashed lines are the bounds of average sea clutter return, and the horizontal dashed line is the visibility factor.

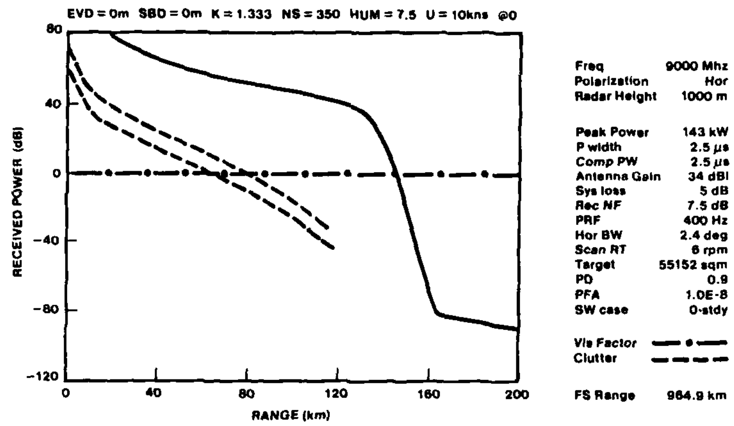


Figure 9. Example Received Power versus Range Display for airborne surface-search radar detection of a naval combatant target.

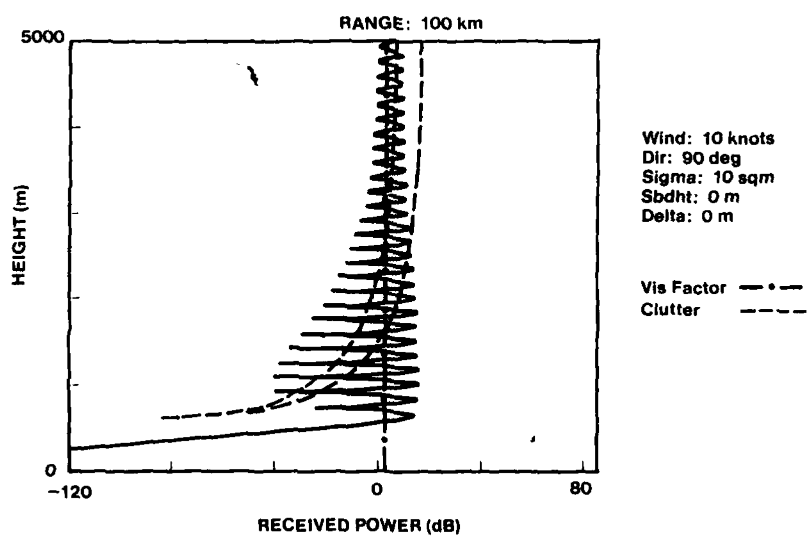


Figure 10. Example display of received power as a function of height for airborne surface-search radar detection of a small surface target.

Engineer's Refractive Effects Prediction System (EREPS)

Herbert V. Hitney
Ocean and Atmospheric Sciences Division
Naval Ocean Systems Center
San Diego, CA 92152-5000

SUMMARY

In recent years, electromagnetic tactical decision aids have been developed to assess environmental effects on the performance of operational systems, such as shipboard radars. In general, these systems have performed well and are now routinely used by operational forces to optimize their use of sensors and deployment of forces. In many cases, laboratory engineers have taken the existing tactical decision aid software and used it to assess the performance of proposed new sensors. Since the original software was not designed for this purpose, many deficiencies in such a use were soon identified. For example, most engineers prefer to graphically compare performance results as a single design parameter, such as radar pulse length, is varied over a range of possible values. Also, in designing a new system, one is usually more interested in the long-term statistical performance than in single-event performance that the tactical decision aids are normally designed to assess.

The Engineer's Refractive Effects Prediction System (EREPS) is a recent development effort tailored to engineering uses and based on the propagation models of the Integrated Refractive Effects Prediction System (IREPS). EREPS is hosted on IBM PC computers for maximum availability to the engineering community, and has been developed using interactive graphics displays for optimum comparison studies. The models are designed in such a way as to give results within a few seconds to allow multiple design trade-off studies to be easily performed. EREPS Revision 1.00 was distributed to interested users in the summer of 1988 and is currently being revised for a summer 1989 distribution. Existing and planned capabilities will be presented along with some examples of applications.

INTRODUCTION

By their very nature, tactical decision aids (TDAs) have been developed for operational users to quickly assess the effects of the natural environment on the performance of systems and to provide displays from which timely tactical decisions can be made. The Integrated Refractive Effects Prediction System (IREPS) was developed in the late 1970's as a TDA for lower atmospheric effects on naval systems in the 100 MHz to 20 GHz frequency range [1,2]. Although it was specifically designed for operational users, IREPS was almost immediately used by some laboratory and industry users to investigate potential propagation effects on systems being considered for development. Since the IREPS design was optimized for single-condition assessment of multiple operational systems, it was soon found to be cumbersome for laboratory simulations, where the effects of multiple simulated environmental conditions or statistical performance were desired. Also cumbersome for this purpose is the system-parameter library, which must be edited through a separate process from the product-generation process. Finally, the IREPS products were designed to support a user, usually an aviation officer, who was different from the IREPS operator, usually an enlisted aerographer's mate. Therefore, interactive-graphics concepts were not used in IREPS. As a result of these considerations, a development was begun in 1986 that has resulted in EREPS, which has been specifically designed to support the scientific or engineering user.

EREPS is a collection of individual IBM-PC programs that have been designed to assist an engineer in properly assessing electromagnetic propagation effects of the lower atmosphere on proposed radar, electronic warfare, or communication systems. Revision 1.00 was distributed in July 1988 [3] and consists of three individual programs named PROPR, SDS, and RAYS. PROPR generates a plot of path loss, propagation factor, or radar signal-to-noise ratio versus range for a variety of environmental conditions from which signal levels relative to a specified threshold or maximum range can be determined. SDS displays an annual historical summary of evaporation duct, surface-based duct, and other meteorological parameters for many 10 by 10 degree squares of the earth's surface. SDS is the primary source of environmental inputs to PROPR. RAYS is a ray-trace program that shows altitude-versus-range trajectories of a series of rays for any user-supplied refractive-index profile, and includes an option to display altitude error relative to a standard atmosphere.

The EREPS programs were developed using Microsoft QuickBASIC Version 3.0 and are distributed as compiled executable files. The basic source code is available by special request. Minimum system requirements are one 360 Kbyte floppy drive, 640 Kbytes of memory, and either a CGA or EGA graphics adapter. As a convenience to potential new users who may acquire copies of EREPS disks indirectly, the user's manual has been included as an ASCII file on the program disks.

CURRENT CAPABILITIES

The PROPR program features three options for displaying propagation effects versus range. The user can select from plots of path loss, propagation factor, or radar signal-to-noise ratio. Path loss is the ratio of transmitted to received powers between isotropic antennas and has widely been used for propagation assessment studies. Propagation factor is the ratio of the actual field strength to the field strength that would occur at the same

range under free-space conditions, and is useful because it can be readily identified in many engineering equations. Signal-to-noise ratio at the radar receiver is a traditional quantity used to assess radar performance. PROPR includes several options for setting performance thresholds, some of which are fully integrated with ESM and radar system parameters, such as receiver sensitivity and transmitter power. For any given type of display, multiple overlays of results dependent on one or more changes in any parameter can be generated, thus allowing easy comparison studies to be performed.

The PROPR models have been designed to operate quickly in a personal computer environment, but without an undue sacrifice in accuracy. In some cases, the models are not as accurate as waveguide or parabolic equation models that have been developed in recent years, but these models are not widely available and generally will not run in an interactive environment on a PC. There are models to account for optical region propagation, diffraction, evaporation ducts, surface-based ducts, tropospheric scatter, water vapor absorption, and sea clutter.

Figure 1 illustrates some of the PROPR capabilities. This case is for a sample shipboard ESM intercept of a shipboard C-band radar. The path loss versus range display option is used, which includes a free-space path loss reference line and an ESM intercept threshold corresponding to a path loss of 166.5 dB, which was computed from the peak power, antenna gain, system loss, and receiver sensitivity shown. Three environmental conditions are overlaid in this figure; a standard atmosphere, a 13-m evaporation duct, and a 300-m surface-based duct. For the standard atmosphere, note the optical, diffraction, and troposcatter regions labeled. Also note the maximum ESM intercept range for a standard atmosphere is about 27 nmi in this case. For the case of the world-average 13-m evaporation duct height, the maximum intercept range is extended to about 60 nmi. For the case of a 300-m surface-based duct, the maximum intercept range is well in excess of 100 nmi, but there is a skip zone or range interval from 27 to 46 nmi in which intercept is not possible. There are no differences indicated in the optical region for the ducting conditions, even though in reality there are effects in this region. PROPR does not currently account for these effects.

Figure 2 shows some PROPR results using the propagation factor versus range display for a sample X-band system. In this case, results are overlaid for evaporation duct heights from 0 to 10 meters in two meter steps. A propagation factor of zero corresponds to free-space conditions. Two additional thresholds have been set at -10 and -20 dB. Note the large variation in propagation factor at each range beyond the horizon attributable to the evaporation duct.

Figure 3 illustrates the radar signal-to-noise display option that includes a sea-clutter model. This example is for a C-band radar located 25 m above the sea versus a 5-m-high target of one square meter radar cross section. The horizontal dashed line is the visibility factor that defines the minimum signal-to-noise ratio that can be detected consistent with the specified probabilities of detection and false alarm (PD and PFA). This visibility factor and the other radar parameters indicated result in a free-space range of 16.4 km. The curved dashed line represents the sea clutter return. The environment in this example is a standard atmosphere with 20 knots of wind speed. The figure indicates the target would be detected at a maximum range of 15 km, but would probably not be detected at ranges less than 6 km due to increasing sea clutter return.

There are several limitations to the current PROPR model that users need be aware of. The frequency limits are 100 MHz to 20 GHz. Only over-water paths are considered and horizontal homogeneity is assumed throughout the program. Antenna heights must be in the range of 1 to 10,000 m, although this limit can be exceeded for one terminal provided the other terminal is low. The evaporation duct and diffraction models are not dependent on the effective-earth radius factor (k) specified, even though some such effects are observed in waveguide models. For a standard value of k=4/3, this limit should not pose a problem. The optical region model does not account for ducting effects, as mentioned previously, and is primarily dependent on the value of k selected. Also, the optical region is truly correct for small elevation angles only. If a user specifies geometries that result in large angles, he should check results against other methods. PROPR uses a single-mode model for the evaporation duct and may be in error for duct heights greater than 30 m at 3 GHz, 22 m at 5 GHz, 14 m at 10 GHz, and 10 m at 18 GHz. Below 2 GHz, PROPR should give acceptable results for all duct heights in the 0 to 40 m range allowed. The surface-based duct model is a single-mode empirical model based on an assumed vertical distribution of refractivity, and is best used to illustrate the skip zone effect and possible range extensions. Finally, the sea clutter model assumes standard atmospheric conditions only, even though it is shown on the same display with target returns that may be enhanced by ducting.

The SDS (Surface Duct Summary) program is used to generate a summary display of long-term annual historical environmental parameters important to near-surface propagation. A convenient feature of this program is the ability to specify a geographic area using a cursor on a world map. One or more of 292 available 10 by 10 degree Marsden Squares can be specified by the user. Once the area is specified, the program produces a numerical and graphical annual summary of the evaporation duct height distribution, average evaporation duct height, and average surface wind speed based on long-term surface observations. Also included are a number of parameters, such as percent occurrence and mean thickness of surface-based ducts, derived from upper-air observations. If a user specifies a single square, and more than one upper-air station exists in that square, there is a procedure to select the desired station for the summary. The data base for the evaporation duct height and surface wind speed was derived from mixed marine surface observations for a 15-year

period compiled by the National Climatic Data Center. The upper-air data was taken from an analysis of 5-years of radiosonde data performed by GTE Sylvania.

Figure 4 illustrates the SDS screen display on which geographical areas are specified. Each of the available 292 Marsden Squares are shown. In this case, six squares in the mid Atlantic Ocean have been selected by moving a "+" cursor and pressing a key to select each square. The resulting summary for this mid-Atlantic area is shown in Figure 5. The evaporation duct height frequency distribution is given in 2 meter increments both numerically and by a bar-chart display. The indicated average evaporation duct height is given as 14.1 m and the average surface wind speed is shown as 13.0 knots. The selected upper-air station is Lajes, Portugal which shows a 10% annual occurrence of surface-based ducts, along with several other parameters derived from the radiosonde data.

The RAYS program generates an altitude-versus-range ray-trace display based on a user-defined refractivity profile of up to 14 linear segments. This profile can be specified in one of several ways. The profile can be entered numerically as either refractivity (N) or modified refractivity (M) versus altitude. The profile can also be defined as N or M versus altitude by moving a cursor on a graphics screen and pressing a special key to define each point. The profile may also be defined by characteristics, such as top and bottom heights of a duct and corresponding trapping layer thickness. Finally, the profile may be specified by radiosonde observations of pressure, temperature, and relative humidity at multiple levels. In addition to the profile specification, the user specifies source altitude, number of rays, and minimum and maximum initial elevation angles. The elevation angles can be set for automatic computation by the program to give a reasonable spread of rays. If the altitude error option is selected, the user must enter the error increment that defines the colors used along each ray to indicate altitude error. The models used in the ray-trace are simply Snell's Law computed using small angle approximations. The altitude error is defined as the difference in altitude between an actual ray and a standard-atmosphere ray of the same initial elevation angle at any point along the actual ray.

Figure 6 illustrates a sample ray-trace for an elevated duct at about 5 km altitude with the source in the duct. The effects of the duct are clearly evident by the trapped rays near 5 km altitude and the radio "hole" at altitudes above the duct. In this case, the profile was specified by characteristics as a duct extending from 4800 to 5200 meters with a trapping layer thickness of 100 m. Figure 7 illustrates a ray-trace generated using the altitude error option. The color along each ray indicates the altitude error for that ray segment based on the color legend shown in the lower right hand corner of the display. In this example, the profile was specified numerically as M-units versus altitude. The case shown is a surface-based duct 1100 feet thick. The antenna height was 100 feet above sea level, representative of a shipboard height-finder radar. Note altitude errors in excess of 4500 meters occur both within the duct and at greater ranges above the duct. These errors can amount to a relative error of more than 20 percent at the greater ranges.

STATISTICAL PERFORMANCE EXAMPLE

This section demonstrates the use of PROPR and SDS to give a statistical assessment of propagation conditions. The complete procedure is detailed in the EREPS user's manual, and is only summarized here. The technique is applicable to situations in which the evaporation duct is the dominant propagation mechanism, such as short-range, near-horizon propagation above 2 GHz, or any surface-to-surface path above 2 GHz in areas where surface-based ducts are rare. The first step is to use PROPR to determine the path loss or detection range versus evaporation duct height for the frequency and geometry of interest. The frequency distribution of duct height is then determined from SDS and combined with the PROPR results to give a frequency distribution of path loss or detection range for the case of interest.

The Greek Island Data of 1972, collected by the Naval Electronics Laboratory Center (a predecessor of the Naval Ocean Systems Center) [4] is used here to compare with statistical assessments made using EREPS. In this experiment, path loss in dB was measured at four frequencies versus time for four three-week periods during 1972 between Mykonos and Naxos, Greece, at a range of 35 km. The transmitters were located about 5 m above sea level and the receivers were located about 19 m above sea level. The four frequencies were 1000 MHz (L band), 3000 MHz (S band), 9600 MHz (X band), and 17500 MHz (Ku band). All of the 1972 data has been combined in the form of accumulated frequency distributions as shown for X band in Figure 8. The dotted curve in this figure shows the percent of all observations that exceed the abscissa value. The dot-dash curve shows the EREPS assessment based on the technique outlined above using PROPR and the annual evaporation duct distribution from SDS for the Marsden Square containing the measurement site. The solid curve shows a similar assessment using the Naval Ocean Systems Center waveguide program named MLAYER, which includes many waveguide modes and should be substantially more accurate than PROPR for higher duct heights. In this case, EREPS and MLAYER are both very close to the measurements below the 70% level, but show substantial differences from the measured data above the 70% level. This discrepancy is thought to be the result of surface-based ducts that are not included in either the EREPS or MLAYER results. At the 50% level, the models and measurements are within 5 dB at a path loss level that is about 25 dB below the diffraction level. Thus it appears that EREPS can be used to give meaningful statistical assessments provided the user is careful not to violate the limits on applicability of the models. Of course, a waveguide or parabolic equation model will always be expected to give better assessments, but such models are not as convenient as EREPS.

FUTURE CAPABILITIES

EREPS revision 2 is currently under development and should be ready for distribution during summer 1989. New products will be an altitude-versus-range coverage display program named COVER, and a path loss, propagation factor, or signal-to-noise ratio plot versus one terminal height to be known as PROPH. PROPH will have all the same general capabilities as PROPR, except that output will be a function of terminal height at a fixed range as opposed to PROPR's output format that is a function of range for a fixed terminal height. In addition to the two new products, revision 2 will have an improved sea-clutter model that has been extended to include effects from the evaporation duct. All of the propagation models will take antenna pattern into account in a manner similar to that used in IREPS. To improve user friendliness, optional support for a mouse will be included that should be most helpful in moving the cursor or cross-hair. Finally, minor bugs from revision 1 will be corrected.

An example of the revision 2 COVER program display is shown in figure 9. This case is for a 425 MHz radar located 20 m above sea level in an environment characterized by a 300-m thick surface-based duct. The two shaded areas on the display correspond to the areas of expected coverage for free-space ranges of 75 and 100 km, which in turn could correspond to two target sizes or two probabilities of detection. Figure 10 is an example of the type of display that PROPH will generate. This example also is for a 300-m thick surface-based duct. A fixed range of 100 km has been selected and the display shows signal-to-noise ratio versus radar altitude from 0 to 5000 m for a fixed target height of 10 m. Note this example includes results from the sea-clutter model, which are shown as two dotted lines that represent the upper and lower bounds of uncertainty about the average clutter level.

The plans for EREPS beyond revision 2 are to improve the propagation models, especially for non-standard conditions. In the optical region, this will result in the use of ray-optics methods that properly take the entire refractivity profile into account. Recent progress in this area is illustrated in figure 11, which compares the results of ray-optics techniques to the Electromagnetic Parabolic Equation (EMPE) program [5] for a 12-m evaporation duct height. The ray-optics results are shown as the solid line, and the EMPE results are shown as a dotted line. Except at the greatest ranges for which ray-optics can be used, the two models are shown to give nearly identical results. The results for a standard atmosphere, also generated by EMPE, are shown in figure 11 as a dashed curve for reference. Note the shifting to the right of the last null in the optical region and the increase in path loss at the last optical maximum. The current models in EREPS would not show these ducting effects at all, since the model assumes a standard atmosphere. The ray-optics results shown in figure 11 were generated in a few seconds on a PC/AT-class computer without a co-processor.

CONCLUSIONS

EREPS has been designed to give fast, easy access to many important propagation effects that are of concern to the system designer. It permits the user to perform easy comparative studies to assess the effects of changes in individual environmental or system parameters. It allows the user to perform some statistical analyses that should be important to system design applications. Although a user must be careful to understand the limits of applicability of the models, EREPS should greatly increase the awareness of propagation effects important to new system design programs. It is anticipated that EREPS will continue to evolve and improve over the next several years, as new displays and capabilities are added and limitations in the models are overcome.

REFERENCES

[1] H.V. Hitney and J.H. Richter, "Integrated Refractive Effects Prediction System (IREPS)," Nav. Eng. J., vol. 88, no. 2, April 1976, pp. 257-262.

[2] W.L. Patterson, C.P. Hattan, H.V. Hitney, R.A. Paulus, K.D. Anderson, and G.E. Lindem, Naval Ocean Systems Center, "IREPS 3.0 User's Manual," September 1987, NOSC TD 1151.

[3] H.V. Hitney, A.E. Barrios, and G.E. Lindem, Naval Ocean Systems Center, "Engineer's Refractive Effects Prediction System (EREPS); Revision 1.00 User's Manual," July 1988, NOSC TD 1342.

[4] J.H. Richter and H.V. Hitney, Naval Ocean Systems Center, "Antenna Heights for the Optimum Utilization of the Evaporation Duct," Vol. 1 & 2, January 1988, NOSC TD 1209.

[5] G. Daniel Dockery, Applied Physics Laboratory, Johns Hopkins University, "Description and Validation of the Electromagnetic Parabolic Equation Propagation Model (EMPE)," September 1987, JHU/APL Report FS-87-152.

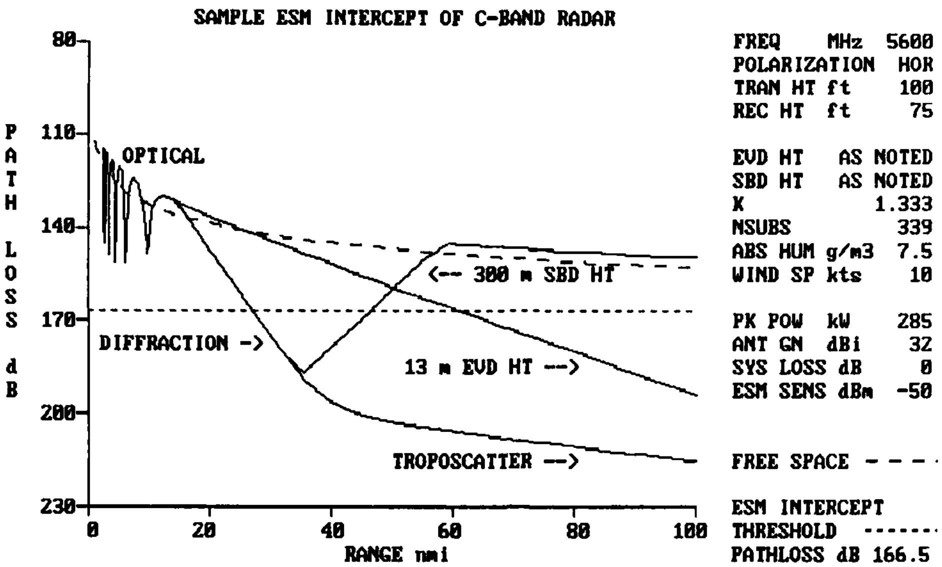


Figure 1. Sample PROPR display for a C-band radar showing intercept ranges for a standard atmosphere, a 13-m evaporation duct, and a 300-m surface-based duct.

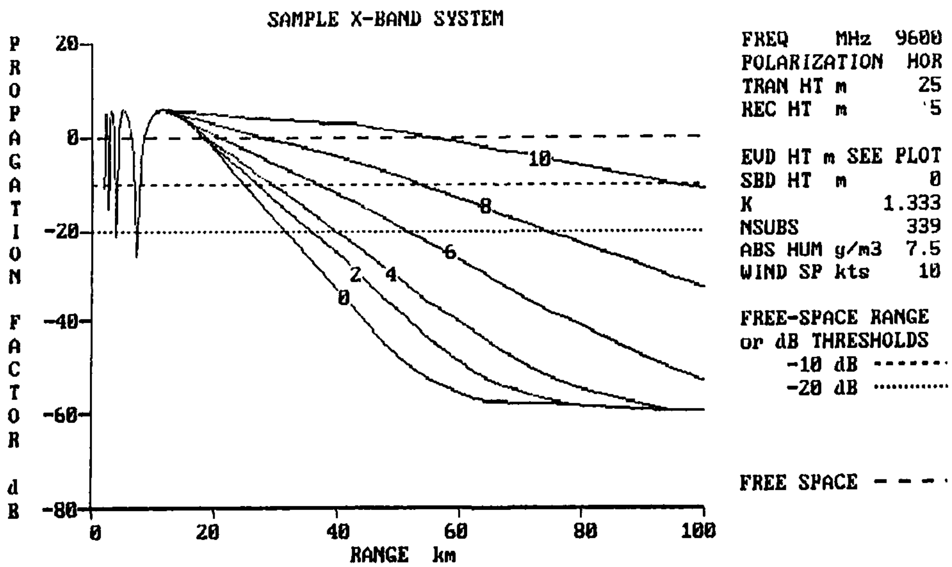


Figure 2. Sample PROPR display for an X-band system showing the effects of 0 through 10 m evaporation duct heights on the propagation factor.

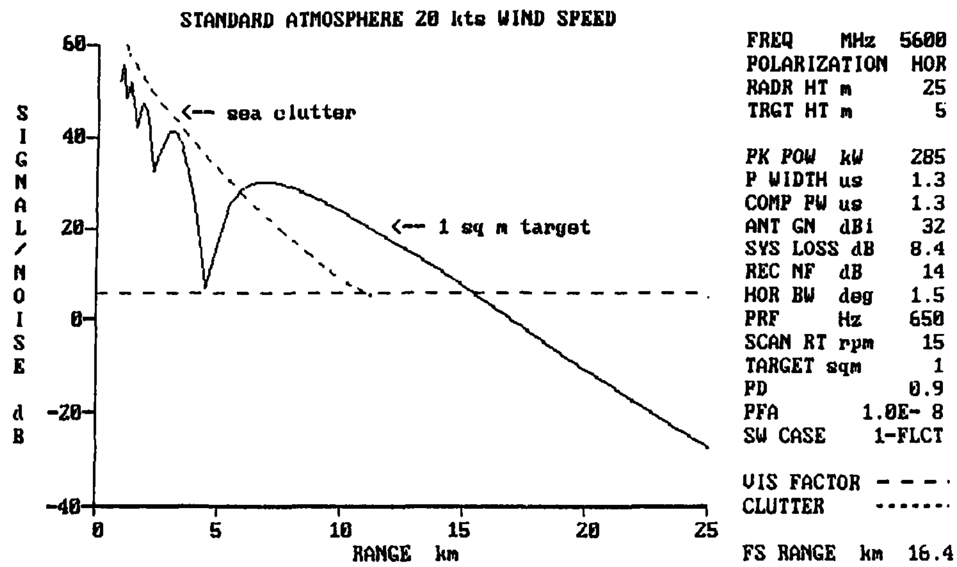


Figure 3. Sample PROPR signal-to-noise ratio display for a standard atmosphere with 20 knots of wind illustrating the return from both a desired target and sea clutter.

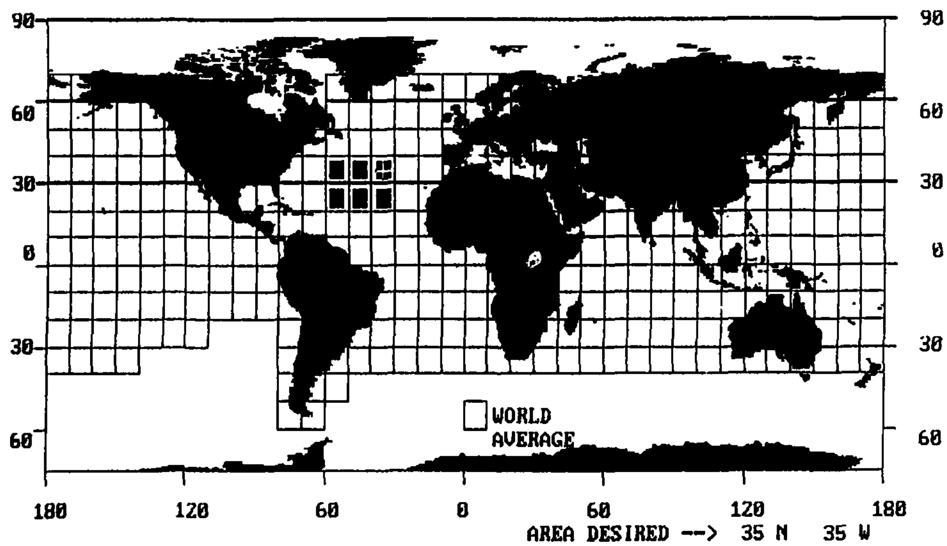


Figure 4. Sample SDS world-map display from which desired geographic areas are selected.

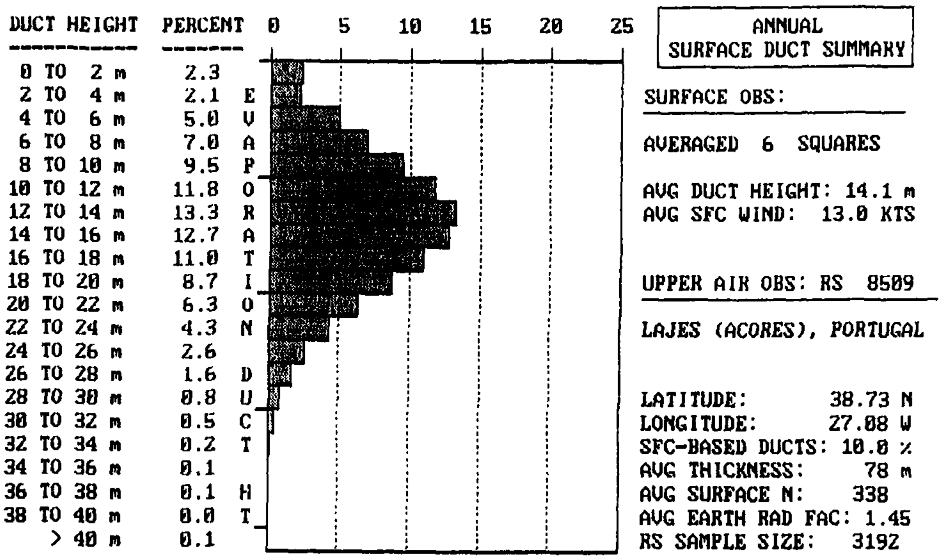


Figure 5. Sample SDS annual surface duct summary display.

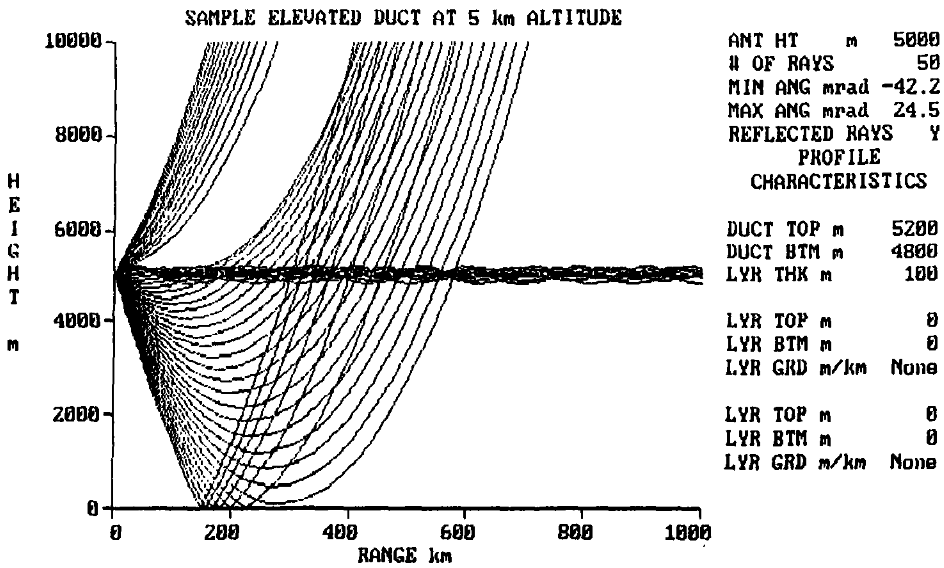


Figure 6. Sample RAYS display for an elevated duct at 5 km altitude with a source in the duct.

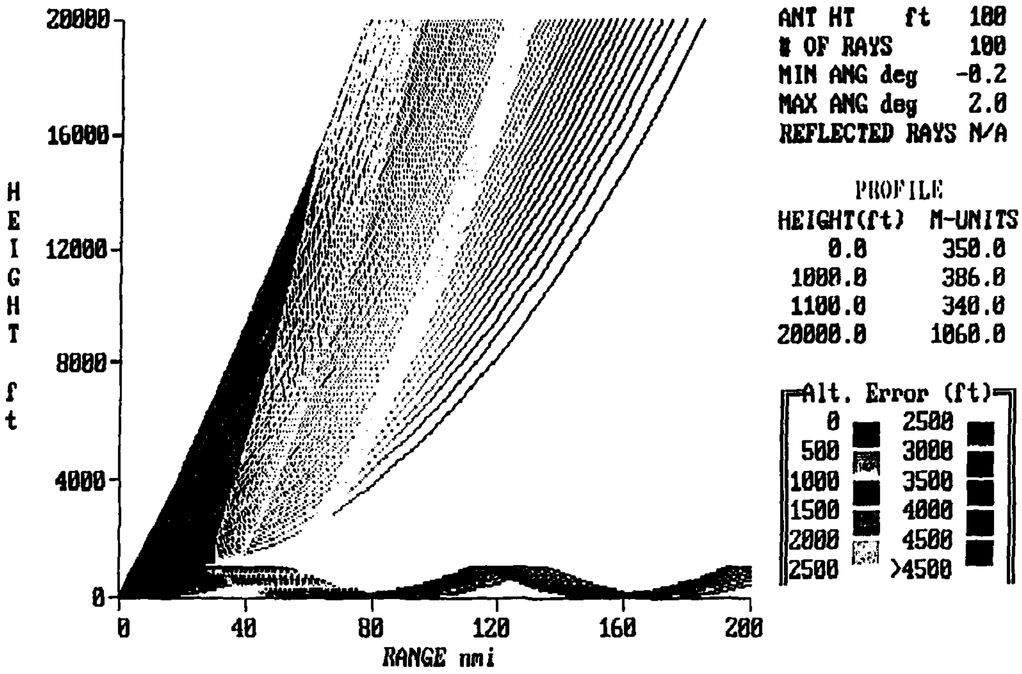


Figure 7. Sample RAYS display for a surface-based duct illustrating the altitude error option.

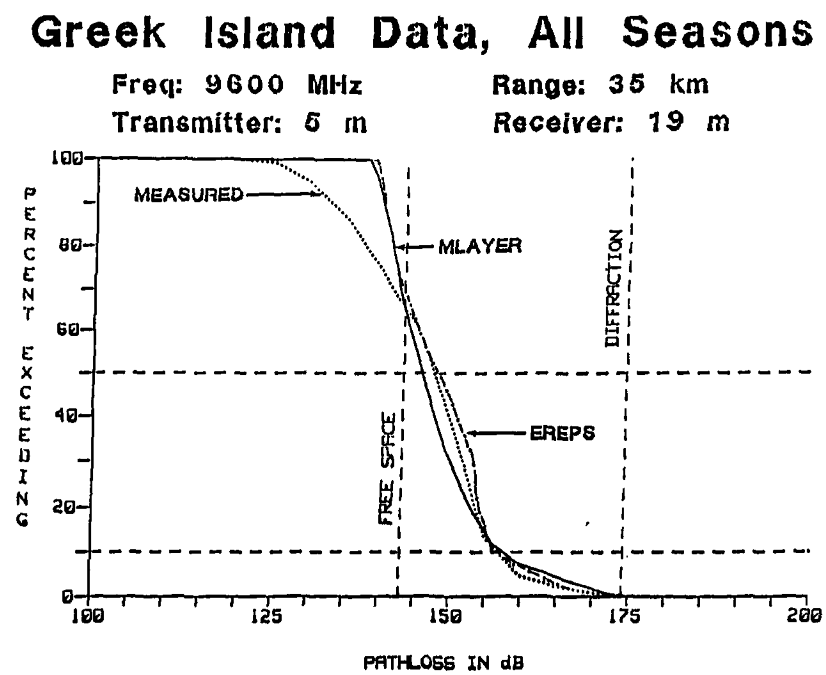


Figure 8. Comparison of Greek Island measurements to results from the EREPS and MLAYER models.

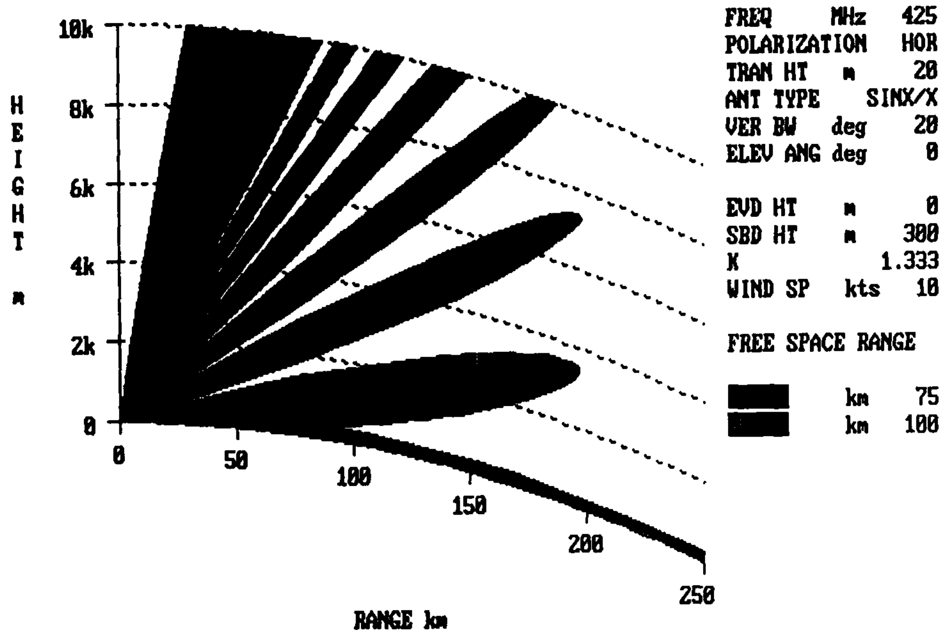


Figure 9. Sample COVER display for a near-surface radar in the presence of a 300-m surface-based duct.

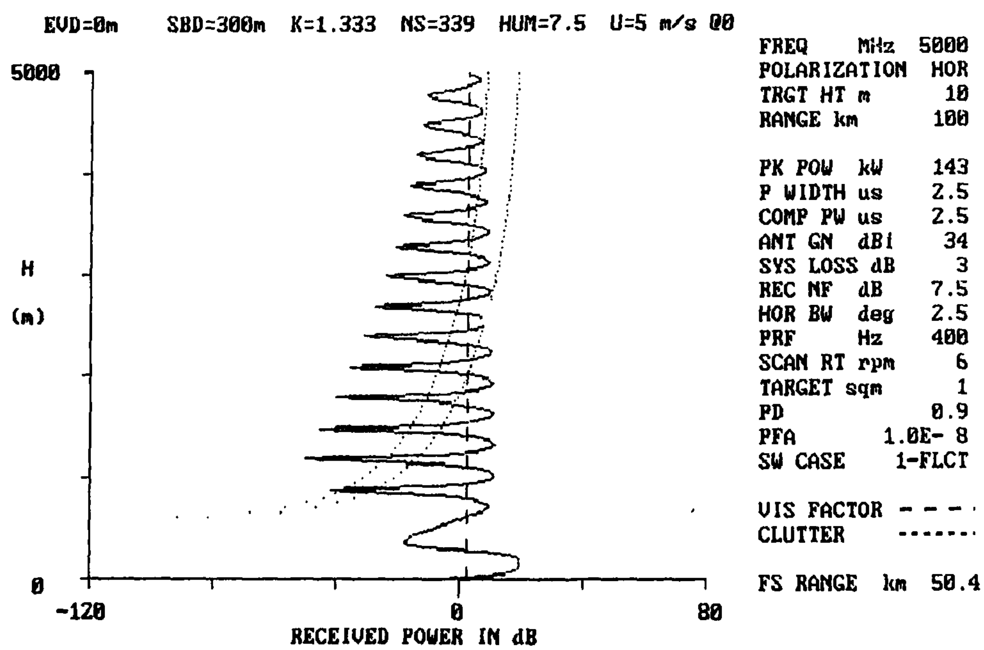


Figure 10. Sample PROPH display for a C-band radar in the presence of a 300-m surface duct showing received power from both a desired target and sea clutter.

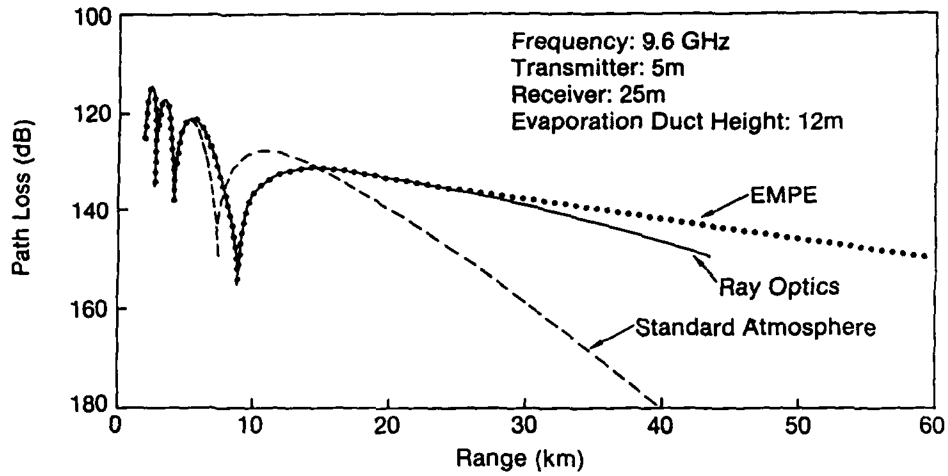


Figure 11. Comparison of ray-optics and EMPE models for a 12-m evaporation duct and a standard atmosphere.

DISCUSSION

R. GOMEZ, US

Have you done a sensitivity analysis of your model to determine which parameters are the principal drivers of your model? That is, how sensitive is your model to the parameters you are using? Are you not doing this work to eventually transition the technology to the operational user? And if that happens, which parameters are going to be used in the operational mode? How good is your prediction as a function of time? How valid is your model?

AUTHOR'S REPLY

EREPS is not intended to eventually transition to the operational user. It is intended to support the scientist or engineer who is developing or evaluating radar, communication, or electronic warfare equipment. Therefore, the emphasis is not on predictions as a function of time, but on average or statistical performance. The sensitivity of the models to input parameters has been checked, and is summarized in "Tropospheric Radio Propagation Assessment," H. V. Hitney, J. H. Richter, R. A. Pappert, K. D. Anderson, and G. B. Baumgartner, Proc. IEEE, Vol. 73, No. 2, Feb 1985.

PROPHET AND FUTURE SIGNAL WARFARE DECISION AIDS

Robert B. Rose
Signals Warfare Division
Naval Ocean Systems Center (NOSC)
San Diego, CA. 92152-5000

SUMMARY

Decision aids, even well designed ones, have demonstrated a wide range of utility and effectiveness when employed in the operational environment. The development and employment of the PROPHET HF (1 -50 MHz.) propagation assessment system has been observed for over a decade. Recent studies indicate that some of the early assumptions on how the module was to be employed were invalid. Although it uniquely fills a real need and its models and products have been extensively tested in operational environments by its users, which include most military and civilian branches of the US Department of Defense, PROPHET faces an uncertain future as a stand-alone system. Its main strength is to provide mission oriented and scenario specific products to perform HF signal coverage, signal vulnerability and radio circuit connectivity in near real time. While technically sound in meeting these objectives, PROPHET's effectiveness is limited by variations in user training, experience, motivation and the lack of operationally convenient access. These problems appear to apply in some degree to the employment of most decision aids. Based on the PROPHET experience, this paper will describe the fallacies that are common in decision aid development and employment, how they affect the decision aid effectiveness, a recommendation for the next generation propagation assessment decision aid and results from an initial demonstration prototyping effort to check out some of the new concepts.

PROPHET: PAST, PRESENT AND FUTURE

PROPHET is the Navy's standard tactical HF prediction decision aid. It is designed to provide real-time information about HF signals warfare and the connectivity of radio communication circuits. Since its inception, PROPHET's development has emphasized operational deployment. Its output products have been designed, to a large extent, at the direct request of its operational users. For example, it's capabilities include the ability to:

(1) predict the band of frequencies that can be used to communicate from one point to another which is not trivial because propagation of signals varies considerably with time of day, season and sunspot cycle;
(2) estimate the expected field strength and signal to noise levels at the receiving end of the circuit;
(3) present a graphical picture of the skywave signal bounce pattern between the earth and the ionosphere;
(4) determine communication frequencies that, due to propagation, are immune to interception or intentional interference.

PROPHET is a self-contained, software system which was conceived in 1974 and first deployed in 1976. It has grown from an initial 5 output products to its present 25 product capabilities, each of which has several sub-options. The most recent release, Version 3.2, has been provided to several hundred users. It uses simple emulations of complex propagation phenomena to take advantage of the portability and speed of personal computers. PROPHET contains large amounts of empirical and intuitive knowledge and user experience, lending it the attributes of an expert system. Throughout its development and employment, the models in PROPHET have undergone extensive verification against "ground truth" oblique-sounder data. Its developers have maintained a continuous dialogue with the user community to ensure continuing operational credibility. This has resulted in a broad spectrum of applications and users.

Today PROPHET's future as a stand-alone capability is in doubt. In spite of its widespread use and usefulness, PROPHET has not been institutionalized or achieved the programmatic acceptance required to sustain, enhance, and promulgate it. Further development and promulgation remain an ad hoc process. This situation prompted PROPHET's developers at NOSC to review the fundamental philosophy of developing and implementing decision aids for the military community. This in-house review indicated errors were made in the assumptions on how the decision aid would be normally employed and identified ways to correct the problems. The remainder of this paper summarizes these findings, proposes recommendations for the next DAs, and describes the results from an initial prototyping effort to check out some of the new concepts.

MAJOR FALLACIES IN DEVELOPING TACTICAL DECISION AIDS

Decision aids (DA), as a generic set of software, typically suffer from at least two fallacies in their concept and development. The first fallacy is that when the user interface is designed, the developer assumes the end user will have the background to ask the right questions in the right order. For example, when each PROPHET module was conceived, it was assumed the user would have the training or experience in HF signals warfare or frequency planning to know the logical inquiry sequence to successfully solve the problem. Review of PROPHET users and the platforms where it has been used indicates that was not a valid assumption. In fact, with respect to HF operations in general, it is more

valid to assume the contrary. In numerous operational tests, it has been observed that PROPHET users span a broad spectrum in motivation and HF experience. The higher the motivation, the more the DA was employed and the more of its capabilities were exploited. In general this motivation was derived from a desire to do the best job possible and from success in using HF. A highly motivated user wanted to exploit PROPHET to its fullest. He knew the propagation medium, the equipment he was using and how to use PROPHET. The HF signals problems were viewed as a challenge to be overcome. Consequently this person's efforts to use PROPHET were successful. This individual soon acquired considerable expertise in using PROPHET and soon became a "key" person in this aspect of normal operations. But eventually this "key" individual would transfer and, often, was replaced by person who could be considered an "antithesis" to the "key" person - a person with little motivation and little or no experience in using the HF medium. If this user, who would rather guess than do any real problem solving, attempted to use PROPHET for frequency planning at all, the attempt was half-hearted with a low probability of success. This self-inflicted bad experience then snowballed, and served as an excuse to not use PROPHET, or any decision aid, again. This, in turn, caused a reduced reliance on the decision aid, further eroding the motivation. When allowed to continue, this degenerative cycle eventually eroded all the progress gained and ultimately the DA was discarded altogether.

The key feature in defeating this negative cycle is to design a user interface that will assure the new decision aid user with the same high level of success, irrespective of motivational level or experience.

The second major fallacy in designing and employing signal assessment decision aids is to assume there will be sufficient time and manpower in the operational environment for them to be used. Generally this is not the case, especially on afloat or airborne platforms. As an example, consider the initial deployment of Classic PROPHET II which was designed to support special direction finding operations. The module was designed specifically to meet the needs of OUTBOARD operators in predicting periods when signals of interest could be expected to be heard. Ships personnel were thoroughly trained in using it on a stand-alone personal computer. However, when the ship got underway, the intended user's time was fully consumed manning their primary positions. There was no extra time to perform the necessary supporting calculations off line, even though the dedicated computer was physically only few feet away. Although the capability was fully functional, its stand-alone location made it impractical for the user to access it.

This experience indicated that a signal assessment decision aid must be embedded into the system it is supporting.

THE NEXT GENERATION DECISION AID

New requirements have been proposed for the next generation of signal assessment decision aid. The proposed design was influenced by the experiences that were just discussed. It must be able to achieve the same high level of performance or productivity regardless of the experience and/or motivation of the user, and must be embedded into the system it is supporting.

The interface of the next generation DA must allow the user to converse with the system in plain language or in the jargon of the specific mission. Expanding telecommunications will make larger amounts of information available to problem solving and the DA must have the ability to peruse these data. Complex scientific calculations, such as those used for propagation and solar-terrestrial relationships will be performed in the computer background, outside the user purview. In many cases, very complicated models, such as those used in PROPHET, will be totally transparent to the user....they will be applied to a users problem in a normal manner, and the result will reflect their use, but their actual operation will not appear in the user system dialogue.

Decision aids, such as PROPHET, are based upon deterministic calculations. The problems amenable to PROPHET for analysis are those for which technical methods provide explicit solutions. However, many other HF/VHF radio operations problems are not amenable to such explicit analysis. An example of this latter situation is the prediction of when and where sporadic-E will occur. Sporadic-E has a profound effect on both HF and VHF systems but can only be described in qualitative terms (e.g.it occurs in May and June at local Sunset and drifts from the southeast to northwest). One of the most promising methods of being able to deal with problems that don't lend themselves to deterministic solution is the Expert System. An Expert System is a computerized implementation of the analytical processes that a human expert employs in his profession. The major elements of an Expert System include the "knowledge base" and the "inference engine". The "knowledge base represents the body of information relevant to the problem under scrutiny and may contain "subjective" rules of the sort used by human experts based on experience; from these rules, the system can construct heuristic solutions. The "inference engine" contains the inference and control strategies that are used to evaluate the knowledge base and synthesize solutions. Expert systems also have the feature of being "knowledge clones" which serves to preserve expertise in various technologies. This is particularly needed in the preservation of HF technology, which has significantly eroded since the middle 1960s.

In 1987/88, developmental work was conducted at the Naval Ocean Systems Center to combine the capabilities of presently-available knowledge-based expert system technology and algorithmic HF propagation-assessment systems such as PROPHET. A module, called the HF Radiolocation System (HFRLS), was developed to test and demonstrate this hybrid concept. The initial software test vehicle was to perform resource management on a hypothetical network of dissimilar HF radiolocation sensors. These included shore-based wide-aperture

HFDF systems, shipboard HFDF systems, single site locator systems and receive-only HF acquisition systems. The purpose of this system was to perform (1) scenario assessment, (2) resource availability and capability assessment, (3) tasking to sensor subnets, and (4) assessment of the decision on whether or not to select a particular sensor. The system was structured so that the operator would start to converse with the system at the start of an event and would monitor progress up through the final solution and output product.

In-house testing of HFRLS indicated that using an "intelligent" man-machine interface greatly enhances the employment of the basic decision aid. The major propagation models in PROPHET were subordinated to the HFRLS system. The knowledge based system "knew" when propagation data was necessary, the right questions to ask PROPHET, and how to format them correctly. This was an internal dialogue that required no user intervention, so the propagation models were completely transparent to the user. The user saw only information needed to perform a radiolocation mission and the DA system's recommendations as to sensor employment. However the final result did reflect consideration of such complex factors as propagation, solar activity, network geometry, and historical sensor performance. The knowledge-based architecture allowed the system to routinely access and peruse large databases for each event. Additional features were included to allow the user to intervene at any level, depending on motivation and experience.

CONCLUSIONS

The survey of PROPHET users indicated that the successful deployment of stand-alone decision aids is vulnerable to the motivation and experience of the user. In addition its employment is heavily influenced by how readily accessible it is to the intended user in his operational environment.

The next generation decision aid will be a hybrid system which will combine knowledge-based expert systems with traditional algorithmic functions such as those in PROPHET. The expert systems will specify, setup, and control complex HF connectivity and signal assessment calculations from PROPHET-like models and carry on the query-response dialogue with the user. To ensure that it is readily accessible operationally, critical issues in the DAs early design are (1) what mission it supports; (2) specifically who the users are and how much time they can devote to the DA operations; and (3) where the DA will be physically located.

The development of the NOSC Advanced HFRLS demonstration terminal provided lessons that are critical in achieving and fully exploiting the next generation decision aid. It was shown that:

(1) Algorithmic signal-assessment systems, such as PROPHET, can be subordinated and tasked by a knowledge based system. This makes the employment of high-level propagation and signal-warfare models transparent to the user.

(2) Subjective information, such as sensor operator experience, sensor performance, and historical trends and characteristics can be employed in the decision process or in the information displayed.

(3) Rule-based expert-system technology allows rapid, intelligent perusal of immense amounts of data which is critical in the control and employment of a large communications or surveillance network. It further provides the basis for "knowledge cloning" which is critical in preserving HF technology.

ACKNOWLEDGEMENT

The author wishes to acknowledge the assistance of Dr. D. R. Lambert in the preparation of this paper.

DISCUSSION

T. DAMBOLDT, GE

The difference between monthly median predicted and measured signal strength can be up to ± 40 dB and the deviation between monthly medians and daily values can be also be up to ± 40 dB. This means that monthly predictions are useless as a decision aid, at least under unfavorable conditions. How do you consider the necessity and value to use real-time solar data to update the model and do you intend to include this feature to the Prophet system?

AUTHOR'S REPLY

Prophet has always contained the ability to input 10.7 cm flux and X-ray flux to provide real time update to the MUF and disturbed LUF calculations. The variation of these values affects the field strength calculation indirectly. Prophet's signal-to-jam ratio calculations which use field strength have routinely demonstrated errors of ± 3-5 dB when interfering HF broadcast signals are measured.

Decision-Aid Design Factors in Connection with HF Communication and Emitter Location Disciplines

John M. Goodman
Naval Research Laboratory
Washington DC 20375-5000

ABSTRACT

The advance of micro-computer technology, the growing sophistication of specified propagation models, and the expanding ability to sense the medium and apply that knowledge in real time is leading to an improvement in the prediction of system performance for tactical users. The maturation of artificial intelligence disciplines should provide the user of advanced C3I decision aids an ability to manage the plethora of information more effectively. This paper identifies critical aspects of the process of developing useful and cost-effective decision aids with emphasis upon the HF medium which is strongly propagation-limited and controlled by variable and often unpredictable phenomena. A major factor in this field of activity is the evolution of "self-adaptive" system architectures incorporating imbedded RTCE. In this context, the decision aid is a process which is operationally transparent to the user but could be user-defined. A key to the development of an adaptive resource management capability is the integration of a set of tools or decision aids which direct the system to compensate for pathological effects by adjustment of system parameters. The approach is ultimately limited by the accuracy with which the ionosphere or the HF channel may be specified. The accepted specification accuracy will determine the design approaches to be followed.

1.0 INTRODUCTION

It is well recognized that the High Frequency (HF) radio regime (3-30 MHz) is the most precarious band in terms of its interaction with the ionosphere. Ionospheric refractivity exhibits considerable variability in both space and time by virtue of its strong dependence upon the ionosphere. Nevertheless it provides for Beyond-Line-Of-Sight (BLOS) communication connectivity and long-range surveillance potential. The nature of HF skywave propagation is known to be strongly influenced by ionospheric personality, and difficulty in the assessment of these traits (or physical properties) may be encountered in support of quasi-adaptive systems which are designed to allow for near-real-time adjustment of (system) parameters ... based upon "trait tracking" and feedback ... to provide for an optimization of overall system performance. The difficulties experienced in this assessment stem from several factors including: the nature of ionospheric variability itself, the mission of the system, and the definition of the system performance metric (i.e., measure of success), which defines the parameters of the assessment process. These parameters involve specification of the most appropriate Real-Time-Channel-Evaluation (RTCE) scheme to be employed, the duty cycle associated with RTCE application, and other notions such as model update, and so on.

With the growth in development of automatic HF radio communication concepts, advanced digital modems, and microprocessor technology, the intrinsic capability to rapidly adjust system parameters to an optimal set has evolved to the point that the ultimate *limit in HF system performance may be* associated with media specification. Media specification is thus the fundamental limit on the capability to tailor performance to the optimal value.

This paper identifies critial aspects in the process of developing useful and cost-effective decision aids in connection with HF communication and HF emitter location disciplines. The first section deals with the systems being considered, and identifies the propagation constraints for each type. The second section discusses the types of decisions to be made, and mentions certain decision aids which are being utilized. The third section concentrates on limitations to the development of resource management tools and decision aids. The concluding section summarizes the paper and provides a set of recommendations.

2.0 SYSTEMS UNDER CONSIDERATION

There are a number of initiatives in the services to address the problems of C2 decision-making. Dacunto [1988] describes current Army C2 initiatives in terms of the Army Command and Control Master Plan (AC2MP), and indicates that the army decision-makers, in order to exercise power effectively, must use information to develop the products of C2. These products are decisions and directives. Information is thus the key. Command and control processing begins with information acquisition and ends with execution of a directive or decision. It is clear that in systems of this type, decision aids will prove to be essential avenues for facilitating the process...and timeliness is normally a desirable attribute.

As indicated at the outset, we shall concentrate on HF systems as a component in the command, control, communication, and Intelligence (C3I) hierarchy. The most obvious HF system type requiring decisions to optimize operation is communication. Others include Over-the-Horizon Radar (OTHR), and Direction Finding (DF) systems. In the latter category, we include various schemes possibly involving measurements of signal parameters at single sites or multiple sites.

The identification of decision aids for "plain vanilla" HF radio systems is relatively simple -- if not obvious -- involving rather traditional approaches for transmitter (frequency) and antenna selection. For adaptive HF communication systems, the situation is less obvious. Nevertheless "Adaptive HF",as distinct from robust designs, requires information concerning the time (and possibly spatially-) varying channel properties.

2.1 HF Communication Systems

The Army has developed battlefield spectrum management schemes to facilitate the function of communication which is termed the "glue" which holds C2 together. Within the HF domain, the Army PROPHET Evaluation System (APES) employs a family of frequency and resource management tools based upon propagation (and possibly threat) models which are resident on a PC. This Army development, and nearly-equivalent versions of the PROPHET methodology addressing needs of the sister services, has brought many of the tools of decision-making to the ultimate user rather than simply the C2 manager. This decentralization may introduce risk if procedures are not properly organized in advance but time from decision to execution is reduced by this form of vertical divestiture. The PROPHET system (and its variants) developed by NOSC is one of the first successful examples of a full class of computer-based decision-aids at HF.

Human operators are the controllers in most operational HF radio systems, but it is evident that modern technological advances are changing the architectures of the next generation systems. Adaptive HF schemes are leading to fundamental improvements in the potential for an improvement in HF communication performance. Automation is the key to effective sytem management and control, and the incorporation of RTCE is an essential ingredient along with the availability of fully-adaptive system components. These components include those which permit changes to be made in selected parameters at various levels: i.e., the transmission level, the link level, the network level, and the overall system level. Figure 1 gives a picture of the decisions which may be involved in the operation of an adaptive HF system.

The role which RTCE plays in the optimization of HF performance has been dealt with extensively in the literature (viz., see Darnell [1986]). It is obvious that RTCE schemes should be imbedded within the HF system itself for maximum effectiveness; nevertheless relatively successful exploitation of non-organic and "out-of-band" ionospheric sounding systems has been well documented. These systems, such as the well-known AN/TRQ-35, provide the system manager with the information which may be used to select the best available frequency for transmission. There are a number of other "systems" which are available for use as sources of RTCE information. Systems which incorporate Automatic-Link-Establishment(ALE) and Automatic-Link-Maintenance (ALM) require that RTCE data be assessed either implicitly or explicitly. The hierarchy of HF ALE systems has described by Harrison [1987]. Harrison identifies ten "logical and unique functions which are critical to automatic linking in the HF environment". The so-called standard levels of automated HF sytems described by Harrison (and depicted in Figure 2) have been characterized at a recent HF information exchange meeting [JTC3A, 1989] as the "stairway to heaven". In the world of HF, each increasing order or level of automation proceeds logically from the lowest adaptability quotient to the highest. Level one corresponds to a "plain vanilla" (non-adaptive) radio and the ultimate system at level ten corresponds to a full complement of adaptive features. It is noteworthy that level 10 functions would be based upon HF-ALE operational rules which would be designed to duplicate the skill levels of experienced HF operators and system managers. Clearly expert system methodologies apply in the realization of Harrison's model of HF system automation.

Artificial Intelligence (AI) schemes have recently been described in connection with the battlefield spectrum management (BSM) problem [Morcerf et al, 1987]. In the BSM arena, automation concepts for system control and operation are driven by increased complexity of the newer and more versatile radios, the need to coexist with a proliferated equipment environment, and a requirement to improve reaction time. On top of this we must also optimize performance in an HF channel which exhibits time-varying capacity. Approaches to HF system design based upon AI disciplines have been discussed by Jowett [1987].

2.2 Emitter Location Systems

High-Frequency-Direction-Finding (HFDF) is the subject of a text by Gething [1978] and abstracts of the early research has been compiled by Tra-

vers and Hixon [1966]. One of the obvious applications of direction finding is emitter geolocation, and the ionosphere is a primary component of error in this process. Large aperture Circularly-Disposed-Array-Antennas (CDAAs) dominate in the world of long-haul HFDF. These systems, which typically measure only the emitter bearing, are operated in a network configuration to provide an estimate of target position by the process of triangulation. Although there may be some problems arising from association of specified signals from netted CDAAs, the system of bearing-only measurement works quite well for emitters located at distances of the order of 500 km or longer.

Large arrays consisting of log-periodic antennas have also been developed. This alternative may be configured to provide elevation and azimuth estimates. Both this scheme and the CDAA are suitable for fixed-site installations.

The US Army fielded a system (the AN/GSQ-85) which had an array dimension of about 150 meters in the decade of the seventies. This array provided both azimuth and elevation information to obtain target position by the so-called Single-Site-Location (SSL) method. Over the years, considerable amount of work has been carried out by Southwest Research Institute (SWRI) in connection with SSL; and systems have also been developed by Sanders Associates and Technology for Communication International (TCI).

The SSL approach has been described by Bennett and Jenkins [1982], Baker and Lambert [1988], Jeffrey et al [1988], and two papers by McNamara [1988a, 1988b].

Other methods have been used to measure emitter location, and a prominent example involves the measurement of the Time-Difference-of-Arrival (TDOA) of the target signal by exploiting accurate time-tagged measurements at three separated sites. This method requires information about ionospheric reflection height. One or more appropriately-located sounders are required to correct for tilt effects which may otherwise vulgarize the results.

Ross [1947] showed the degradation in bearing angle as a function of distance to the emitter. The bearing errors are quite striking at the higher elevation angles and once again ionospheric tilts are the designated culprits. There is a tradeoff between antenna aperture and (emitter) signal averaging time, with greater averaging necessitated if the aperture is small.

There is a definite requirement for location of HF emitters in the tactical environment for which source distances may be less than 500 km. As indicated in the previous paragraph, tilts may introduce a relatively large azimuthal error in this regime which is sometimes referred to as the Near-Vertical-Incidence-Skywave (NVIS) mode of propagation. This mode presents a technical challenge for HFDF. The SSL method is often considered in the quest for a solution to the geolocation problem associated with NVIS signals. Even so, the SSL method also experiences a difficulty with overhead tilts. This is not only a matter of inexact ionospheric specification, but ray propagation as well.

3.0 TYPES OF DECISIONS TO BE MADE

The classes of decisions associated with HF systems may be conveniently broken into levels all of which are consistent with the overall requirement involved. The more adaptive we desire to make a system, the more we must strive to place real-time-decision (RTD) capability at the lowest possible levels.

Real-Time-Channel-Evaluation (RTCE) systems aid in this process and RTCE is actually a sub-category of a more general class of capabilities which we shall refer to as Real-Time-Decision-Aids (RTDA).

We shall define RTDA to be a class of functional capabilities or systems which allow for decisions to be made at the specified system levels and at designated times. While highest-level system-wide decisions may be made in real-time, they are more likely to be required in the operational planning phase where the greatest lead times are involved. The lowest-level decisions, on the other hand, may not be made in advance. These decisions place the greatest stress on the performance of the system in terms of speed and accuracy. Automation is clearly the key at this basic level, and human intervention must be minimized.

The decision tree may be viewed as a set of discrete levels even though it is more probably a continuum (involving leaves as well as branches).

3.1 HF Communication Decision Aids

It is worth remembering that the nature of the ionosphere may introduce major changes in the level (or order) of decision. A major propagation disturbance, for example, may necessitate a system-wide decision. In the context of communication, a switch in media (say from HF to MF groundwave, meteor burst,

or even satellite) may be required. Possibly an HF network decision may suffice to circumvent a disturbance. Another strategy is to relax the adaptivity requirement and opt for a class of robust procedures. This class of schemes has been developed over the years in the context of the nuclear environment for which message integrity is essential and high data rates are not. In the example just presented, we find that media diversity is specified up-front in a robust design, and there is no decision to be made. In short, for robust designs, we opt for reduction in throughput (but at greatly reduced risk) ;but for adaptive designs, we engineer the system for maximum capacity (but with greatly increased risk). The risk factor is related to the performance of the RTDA's which may be incorporated into the adaptive system.

As we have suggested, information type and rate naturally depends upon the level of decision to be made. Starting at the lowest level, information detailing the channel at (fixed frequency) is required but this information must be obtained and exploited at a speed which is faster than the time constant of the ionosphere. In the parlance of adaptive HF, the so-called learning constraint must be satisfied. The decision aid in this case is a training sequence which may precede the traffic or be imbedded in the waveform itself. The decision (which is really an ionospheric compensation) is made by an intelligent modem. The decision process is dependent upon another condition being satisfied; namely the diversity constraint.

At the next level up, we have decisions which involve link-level operations.

At an intermediate level, decision aids are less sensitive to details of ionospheric structure and motion. Accordingly, the temporal and spatial resolution associated with media sampling (via RTCE) is relaxed. Sounding systems may be utilized at this level to specify the general amount of spectrum space available (i.e., the LOF to MOF band).

Other classes of RTCE may also be exploited at this level, and they include various schemes, a popular one being a variation of the pilot-tone concept.

Sounder-updating is a method of major interest and the subject has most recently been described by Goodman and Daehler [1988]. Some models are more useful in the update mode than others, and this matter has been investigated by Reilly and Daehler [1986]. It must be remembered that the update method is principally of interest in the context of the non-cooperative link or when handshaking is not advisable or possible. Additionally, it is obvious that models which are updated by ionospheric probe/sounder data provide for no prediction improvement if the model is exercized at the "control point" associated with the enabling update. By the same token, ionospheric sounder data loses considerable value if it is exploited in an environment for which the communication path and sounder path "control points" are widely separated. The marriage of the sounder data with an appropriate model therefore provides for a degree of performance improvement in the denied area situation.

3.2 Emitter Location Decision Aids

In the HFDF arena, ionospheric impact is greatest when using the SSL method. Again, we may consider decision aids at the highest level down to the lowest. At the highest level, the planning aspects predominate, and models may be exploited to ascertain the intercept probability of various classes of signals under median ionospheric conditions. This problem is similar in principle to the climatological prediction of jamming threat.

RTCE data may also be used to advantage at lower levels in SSL system management and operation. In general, we may obtain information from the following RTCE systems:

a. Ensemble of cooperative HF emitters serving as check targets

b. Sounders (vertical and oblique)

c. Over-the-Horizon (OTH) radar

d. Satellite-borne sensors to sample denied areas (i.e.,topside sounders and UV imaging systems)

e. The target emitter itself

Since a and b above are associated with paths which will not be coincident (in general) with the wanted signal, we will be dealing with a situation in which the ionospheric assessment may be uncorrelated with the path for which ionospheric information is required. On the other hand, c, d, and e contain information about the path of interest. OTH radars and satellite

systems are expensive and the incorporation of such systems for the sole purpose of ionospheric compensation for HFDF would be unaffordable. The process by which propagation data may be extracted from the target emitter signal pattern for ultimate exploitation in ionospheric compensation is complex; suffice it to say that category e must be combined with one or more additional RTCE schemes.

Sounder update methods would appear to have the greatest application in the context of HFDF.

4.0 LIMITATIONS ON SYSTEM PERFORMANCE IMPOSED BY THE IONOSPHERE

The fundamental limits are imposed by two factors which are difficult to separate: propagation (magneto-ionic properties, anisotropy, dispersion) and medium (including ionospheric structure and dynamics).

For HF communication ,structural details define the impulse response and the dynamical behavior is central in development of the scattering function. This information may be obtained from wideband probes [Wagner and Goldstein, 1985]. Models of the HF channel have been obtained for both the narrowband [Watterson et al, 1979] and the wideband cases [Nesenbergs, 1987; Vogler and Hoffmeyer, 1988]. Nevertheless these models are not capable of providing the information required for ionospheric compensation in real time.

HF communication systems differ in many respects, so it is not surprising that the media constraints are system-dependent as well. Analog systems suffer from fading problems, and digital systems encounter intersymbol interference as well. Wideband systems have advantages which are well known. Diadvantages are associated with the dispersive nature of the ionosphere. Modern techniques (RAKE with parallel processors and adaptive equalization) may be used to achieve implicit diversity...thereby compensating for fading effects to a large degree, and intersymbol interference (ISI) may be resolved as well. Also such measures allow for an increased data rate which is normally restricted by multipath spread. But there are still diversity and learning constraints which must be satisfied.

Ray tracing methods are finding application, especially in the context of HFDF and OTH radar. Virtual "raytracing" methods are still used in mainframe models such as IONCAP but 3-D raytracing is re-emerging in special applications. This is significant because with the possibility (indeed the necessity) of improved ionospheric specification, a far greater burden will be placed upon the raytracing component in a prediction/assessment algorithm. 3-D raytracing methods are now being applied in instances which require an accurate relation between known launch angles at one end of the ray path and geolocation at the other end. A well-known 3-D raytracing program, based upon numerical solution of the Hamiltonian or Haselgrove equations was developed by Jones and Stephenson [1975]. Recently, Reilly and Stroble [1988] of NRL have developed a 3-D raytracing code based upon an analytic solution of the Euler-Lagrange equations for each raypath increment. These workers also combined the code with a climatological model [Thomason et al, 1979].

5.CONCLUSION

In conclusion, we have shown that decision aids come in many forms and address a variety of problem areas. Ionospheric specification is the central ingredient for which algorithms are required in specified decision aids. Future aids will incorporate RTCE to a greater extent than found in current systems. Improvement in ionospheric (and equivalently HF channel) specification brought about by RTCE incorporation, and the transformation of derived information into parameters which will allow for ionospheric compensation is urgently required for any significant improvement in HF system performance to be realized. In addition to an improved specification of the ionosphere, it is essential that state-of-the-art 3D raytracing techniques (including magnetic field effects) be incorporated into the codes. This statement is most appropriate for HFDF applications.

Future decision aids must provide timely assessment of local and remote tilts as well as their predicted behavior. They must provide for a timely and accurate specification of the position of circumpolar features such as auroral arcs, the midlatitude trough, and the polar cap. Also, for operation in the trans-equatorial environment, a precise assessment of the Appleton anomaly is needed.

Currently there are a number of initiatives directed toward the improvement of global ionospheric models, and work is also being undertaken in the area of HF propagation modelling. NRL is continuing the development of an oblique-sounder data base for use by the scientific community, although the progress has been slowed due to a reduction in support. Several experimental data collection campaigns have been initiated in the last few years and an international program to monitor calibrated HF transmissions has been proposed. Unfortunately this activity is only loosely coupled with the develop-

ment (and deployment) of a variety of RTCE devices. Coordination of the various model studies and RTCE experimentation is urged to allow for the development of more efficient RTDAs and decision aids in general.

Finally, it should be recognized that existing (main frame) physical models of ionospheric behavior are not suited to provide the detailed real-time information required in advanced HF systems. Recognized HF propagation models such as IONCAP are no better in this context. Both classes of models improve considerably when specified parameters are "updated". This is especially true of F region parameters which are principally driven by atmospheric/ionospheric dynamical and diffusion processes. These processes resemble a type of chaos which is not amenable to prediction. Fortunately the ionospheric "state" -- once specified -- does obey certain general rules in terms of its spatial and temporal correlation. As a result of this property, model update approaches have found application in (a) the forecasting of system performance over uncooperative links, and (b) ray tracing within denied areas. Models which are used in this way may be decidedly inelegant. Accordingly, simple PC-based models will suffice for most real-time applications (i.e., RTDA). The large physical models have their greatest application in the long-term prediction process, and if coupled with a suitable propagation sub-model, compare well with climatological propagation models such as IONCAP, RADARC, and AMBCOM. By far the greatest value which will accrue from development of theoretical (physical) models is the insight they provide. They are well suited for sensitivity analyses, while other models are not optimized for that purpose.

REFERENCES

Baker D.C. and S. Lambert, 1988, "Range Estimation for SSL HFDF Systems by Means of a Multi-Quasiparabolic Ionospheric Model", private communication to JMG in anticipation of publication in IEE Proceedings, Part H, Micro-waves, Antennas and Propagation.

Bennett S.M. and K.B. Jenkins, 1982, "Advanced HFDF System Provides Ranging by Vertical Triangulation", Defense Electronics, June issue.

Dacunto L.J.,1988, "Army Command and Control Initiatives", Signal, November, pp. 63-69.

Darnell M., 1986, "Imbedded Real-Time-Channel-Evaluation", in NATO-AGARD-LS-146, Propagation Impact on Modern HF Communication System Design.

Gething P.J.D.,1978, Radio Direction Finding and the Resolution of Multicomponent Wave-fields, published by Peter Perigrinus Ltd(IEE),Southgate House, Stevenage, England.

Goodman J.M., and M.H. Reilly, "Shortwave Propagation Prediction Methodologies", IEEE Trans.Broadcasting, Vol.34, No.2, pp 230-237.

Harrison G.L., 1987, "HF Radio Automatic Link Establishment Systems", IEEE Conference, (2.1.1-2.1.5.), pp. 43-47.

Jeffrey Z.R., P.T.Middleton, and C. Winkler, 1988, "Accurate Measurements of the Total Angle of Arrival of HF Radiowaves", in NATO-AGARD-CP, Munich Meeting on Ionospheric Structure and Variability on a Global Scale and Interaction with Atmosphere and Magnetosphere.

Jones R.M., and J.J.Stephenson, 1975, "A Three Dimensional Ray Tracing Computer Program for Radio Waves in the Ionosphere", OT Report 75-76 (PB 248856) Dept. of Commerce, Boulder CO.

Jowett A.P., 1987, "Artificial Intelligence in HF Communication Systems", in NATO-AGARD-CPP-420, Lisbon Conference, Paper No. 14., pp. 14-1 to 14-25.

JTC3A, 1989, HF Informatior Exchange Meeting, held at BDM, 20-23 March

McNamara L.F., 1988, "Ionospheric Modelling in Support of Single Station Location of Long Range Transmitters", J.Atmospheric Terrest. Phys.

McNamara L.F., 1988, "Ionospheric Limitations to the Accuracy of SSL Estimates of HF Transmitter Locations", in NATO-AGARD-CP, Munich Meeting on Ionospheric Structure and Variability on a Global Scale and Interaction with Atmosphere and Magnetosphere.

Morcerf L.A., K.R. Kontson, and J.E. Clema, 1987, "A Concept for the Application of Artificial Intelligence Technology to Battlefield Spectrum Management", IEEE Conference, 5.1.1-5.1.6

Nesenbergs M., 1988,"Modeling of Wideband HF Channels",in Scattering and Propagation in Random Media (AGARD-CP-419), edited by K.C. Yeh and A.N. Ince, Specialised Printing Services Ltd.,Loughton Essex UK.

Reilly M.H., and M. Daehler, 1986, "Sounder Updates for Statistical Model Predictions of Maximum Usable Frequencies on HF Skywave Paths", Radio Science, Vol 21, No. 6, pp. 1001-1008.

Reilly M.H., and E.L. Stroble, 1988, "Efficient Ray Tracing Through a Realistic Ionosphere", in Effect of the Ionosphere on Communication, Navigation, and Surveillance Systems (IES'87) edited by J.M. Goodman, US Govt. Printing Office, Washington DC

Ross W., 1947,"Estimation of the Probable Accuracy of High Frequency Radio Direction Finding Bearings",J.IEE,Vol.94,Part IIIA,No.15,pp.722-726.

Thomason J., G. Skaggs, and J. Lloyd, 1979], "A Global Ionospheric Model", NRL Memo Report 8321, Naval Research Laboratory, Washington DC 20375-5000.

Travers D.N. and S.M. Hixon, 1965, Abstracts of the Available Literature on Radio Direction Finding 1899-1965, Southwest Research Institute, San Antonio TX.

Vogler L.E. and J.A. Hoffmeyer, 1988,"A New Approach to HF Channel Modeling and Simulation: Part 1:Deterministic Model",NTIA Report 88-240, Dept of Commerce, Boulder CO.

Wagner L.S. and J. Goldstein, 1985,"High Resolution Probing of the HF Ionospheric Skywave Channel:F2 Layer Results",Radio Science, 20(3),pp.287-302.

Watterson C.C., J.R.Juroshek, and W.D. Bensema,1970, "Experimental Confirmation of an HF Channel Model",IEEE Trans.Comm.Tech.(COM-18), pp.792-803

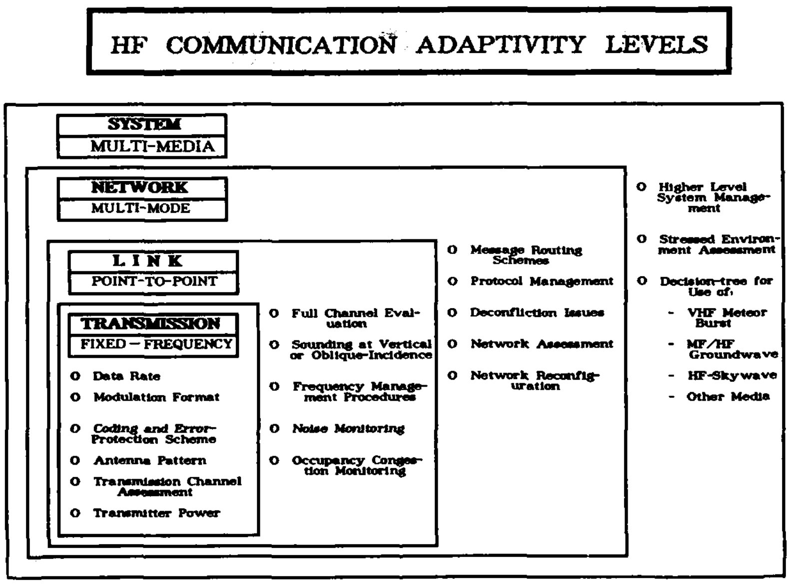
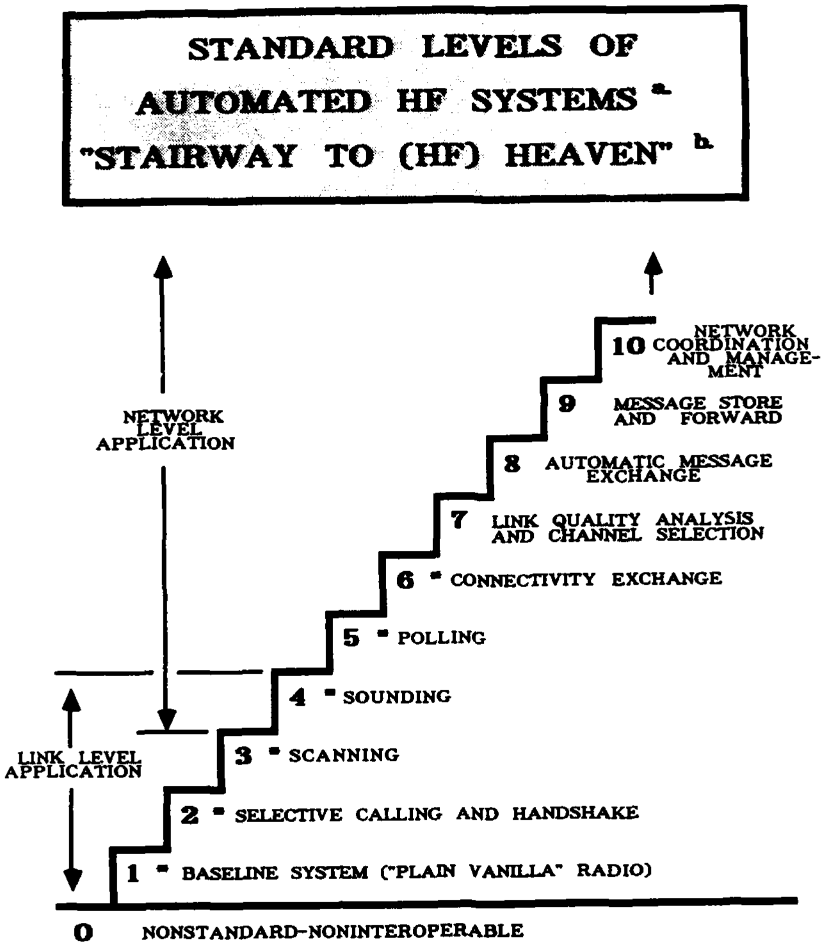


FIGURE 1. The four basic levels of HF communication adaptivity (BOLD FACE CAPS). Also given is a one-word description (CAPS) and a set of attributes for each level.



a. From HARRISON, G., 1987, "HF Automatic Link Establishment Systems", in IEEE Proceedings, reproduced by permission of author.

b. Unknown source, 1989 HF Information Exchange Meeting, McLean, VA.

FIGURE 2. The ten levels of HF automation.

DISCUSSION

D. YAVUZ, TUR

Could you comment on the following observation: the HF band is densely occupied, i.e. it has on the average about two orders of magnitude higher radiated electromagnetic energy per Hz (or KHz etc.) than other bands. This results in very high levels of random interference. Thus even if we had a perfect ionospheric propagation model as a decision aid, this would not imply a good communications performance. Is there not, therefore, a danger in "overrefining" HF propagation models at the expense of efforts to improve link/network layer protocols which would provide much more cost effective improvements in overall communications performance/service.

AUTHOR'S REPLY

Yes, there is always such a danger in situations as complex as HF system design and operation. Certainly even a "perfect HF propagation model" is not the only decision-aid" we must consider. As you point out, other "models" from which network and link level decisions are based must also be included in the mix. Efforts to improve link/network layer protocols should, I feel, proceed in parallel with HF propagation model improvement. I also suspect that better HF propagation models (properly updated using appropriate real-time channel evaluation methods) will reduce to some degree the complexity of protocol development and application. Moreover, better propagation assessments have unique applications in the ECM arena and in non-cooperative HF communications.

Long Wave Propagation Assessment

Jerry A. Ferguson
F. Perry Snyder
Ocean and Atmospheric Sciences Division
Naval Ocean Systems Center
San Diego, CA 92152-5000
USA
(619) 553-3062

SUMMARY

This paper describes the long wavelength propagation capability of the Naval Ocean Systems Center. This capability applies the concept of the earth-ionosphere waveguide to very low frequency (vlf: 10 kHz to 30 kHz) and low frequency (lf: 30 kHz to 100 kHz) radio propagation. It is used to perform assessment of communications coverage at long wavelengths via graphical displays of signal strength and signal to noise over individual propagation paths and over large geographical areas.

The propagation assessment capability is a collection of computer programs which are exercised separately or in sequence depending on the application. The execution of these individual programs for purposes of obtaining coverage maps can be set up by a driver program which automates the process. At the present time, this driver program implements the sequencing of the programs via Digital Equipment's VAX VMS control language.

The automation of the calculations depends on accurate specification of the parameters of the earth-ionosphere waveguide. The driver program contains models for these parameters which have been developed over a number of years. The model of the ionosphere is an approximate one which characterizes the conductivity as an exponentially increasing function of height. Some allowance for the nighttime differences between temperate and polar latitudes is made. The lower boundary of the waveguide is defined by a detailed ground conductivity map based on geological formations.

INTRODUCTION

The model of longwave propagation developed at the Naval Ocean Systems Center is based on the concept of the earth-ionosphere waveguide (Budden, 1961; Pappert, Gossard and Rothmuller, 1967). Signal levels are found by obtaining modal solutions to the specific waveguide under consideration. Complicated propagation paths are treated as a series of horizontally homogeneous segments. The parameters of the segments are determined by the local values of ground conductivity, the magnitude and orientation of the gemagnetic field and the state of the ionosphere.

A widely used computer program which makes the necessary calculations for a single horizontally homogeneous earth-ionosphere waveguide is "MODESRCH" (Morfitt and Shellman, 1976). An improved version of MODESRCH", named "MODEFNDR", has recently been developed (Shellman, 1986). This latter program is an important component of the propagation capability and will be denoted by "MF". Introduction of varying earth-ionosphere waveguide parameters in this model is processed by the Segmented Waveguide program, denoted as "SW", (Ferguson and Snyder, 1986). It requires as input an initial set of waveguide parameters and associated mode solutions. The initial waveguide parameters need not be at the transmitter. The program then extrapolates the solutions as the waveguide parameters vary with distance from the transmitter.

The mode solutions calculated along a propagation path are ultimately used to obtain signal strength. After the mode solutions have been obtained, the strength of the electromagnetic field along a path can be determined using one of two mode conversion models, one denoted "FULLMC" (Pappert and Shockey, 1972) and the other denoted "FASTMC" (Ferguson and Snyder, 1980). "FULLMC" does full wave calculations in the process of calculating mode conversion coefficients and can be quite slow in execution time; whereas, "FASTMC" is an approximate model and runs quite quickly. The two programs generally produce comparable results (Pappert and Ferguson, 1986).

After defining the propagation path (direction, length, frequency, etc.), the process of obtaining accurate solutions to the earth-ionosphere waveguide along the path involves determining the proper path segmentation by consideration of the variation of the ionospheric profiles, orientation of the geomagnetic field and the ground conductivity. The solutions can be quite sensitive to small changes in parameters of the earth-ionosphere waveguide. In addition, paths in the earth-ionosphere waveguide can experience large changes in ground conductivity such as are encountered in going from sea water to thick ice caps (Greenland). Because the mode solutions are dramatically different between sea water and ice, it is not possible to determine the mode solutions for one ground conductivity using the solutions for the other. Therefore, the extrapolation procedure used by "SW" fails across such changes.

A driver program has been developed to automate the execution of programs and maintenance of the files required to obtain the final output. It is also designed to mitigate some of the problems described above. It sets up files for input to and execution of the basic programs "MF" and "SW". The outputs from these runs are then concatenated for use in either of the aforementioned mode summing programs. However, the driver program currently sets up execution of the approximate mode summing model.

The current implementation of the propagation capability is being developed on a Digital Equipment VAX computer operating under the VMS operating system. The command files which are created by "PRESEG" contain instructions for establishing temporary files which are used by a procedure written in VMS. It sets up data files for "MF" to obtain starting mode solutions for a specific segment on the propagatin path. These

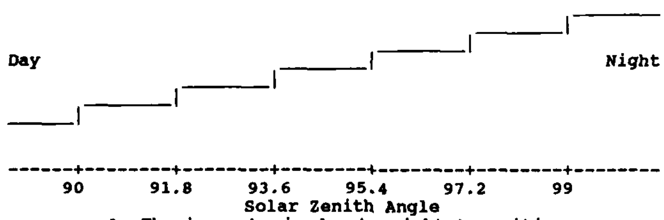
starting solutions are input to the program "SW" which attempts to obtain solutions to a series of homogeneous segments along the propagation path using an extrapolation procedure. This procedure may fail to reach the end of the propagation path for a number of reasons. The status of the execution along the path is checked by the VMS driver. Depending on this status, the procedure sets up another execution of "MF" in order to continue execution of "SW" at the current segment or it skips the current segment and continues with the next one.

GEOPHYSICAL MODEL

Geometrically, a propagation path is a great circle on the surface of the earth which passes through the transmitter. The path setup program determines the ground conductivity and dielectric constant, the orientation of the geomagnetic field and the solar zenith angle at the transmitter and at successive points along the path separated by small, fixed distance intervals, . The ground conductivity model is derived from that of Morgan (1968). This model has 10 levels of conductivity with 0.5° resolution in latitude and longitude.

The solar zenith angle is used to specify the ionospheric profile at each path point. For nighttime paths, the geomagnetic dip angle is also used to specify the ionospheric profile. There are two transitions in the ionosphere. One of these is from day to night. The other transition occurs along paths which are in night that pass from low and middle geomagnetic latitudes into or out of polar latitudes. The daytime ionosphere is specified for solar zenith angles less than 90° and the nighttime ionosphere for solar zenith angles greater than 99°. The nighttime latitudinal transition from middle to polar latitudes takes place between geomagnetic dip angles of 70° and 74°, although the user has control over this range of values.

The ionospheric profile model is defined by an exponential increase in conductivity with height specified by a slope in km^{-1} and a reference height, h', in km. A value for the slope and h' is required from the user for both daytime and nighttime at two frequencies. Linear interpolation is used to obtain two values each of slope and h', one for day and one for night, at the user specified frequency. These two values are used to define 5 horizontally homogenous segments. Each segment is characterized by a slope and h' which is obtained by linear interpolation between the day and night values. These 5 segments define the basic dawn/dusk transition. When the path is fully night, h' also depends on the geomagnetic dip angle. This dependence is chosen so that the h' for the polar nighttime ionosphere is the midpoint of the intervals between the daytime and nighttime as illustrated in Figure 1b. These polar values will be denoted by the subscript 'P'. This crude model is used in order to develop a reasonable set of ionospheric profiles to handle all of the transitions. A more sophisticated model is not warranted due to a lack of data. However, the parameters are easily modified by the user for special studies.



A: The ionospheric day to night transition.

B: The ionospheric polar cap transition at night.

Figure 1: Illustration of the ionospheric transitions.

A useful model of the ionospheric profile can derived from Morfitt (1977), Ferguson (1980) and Morfitt, Ferguson and Snyder (1981). This model defines the daytime ionosphere by a slope of 0.3 and an h' of 74. The nighttime ionosphere is more complicated in that the slope varies linearly with frequency from 0.3 at 10 kHz to 0.8 at 60 kHz. The low- and mid-geomagnetic latitude nighttime ionosphere is characterized by an h' of 87 while the polar nighttime ionosphere has an h' of 80 km. Values of these

transition parameters at 30 kHz are found in Table 1. Figure 2 illustrates the two transitions as they would be defined along a hypothetical path which traverses the pole from day to night.

Table 1: Transition Parameters

Solar Zenith Angle	slope	h'	Magnetic Dip Angle
X<90 (Day)	0.30	74.0	D<70
90 <X<91.8	0.33	76.2	70<D<72
91.8<X<93.6	0.37	78.3	72<D<74
93.6<X<95.4	0.40	80.5	74<D<90 (Pole)
95.4<X<97.2	0.43	82.7	72<D<74
97.2<X<99	0.47	84.8	70<D<72
99 <X (Night)	0.50	87.0	D<70

```
                                                             |________
Day                           Polar Cap       ____|_____|    Night
                                    _________|
                     _______|
          ______|
_____|

h'=74    76.2     78.3       80.5        82.7  84.8   87

_____|_______|_______|_________|____|____|________
  X=90     91.8    93.6       D=74   72   70
```

Figure 2: A hypothetical trans-polar path which crosses the dawn/dusk line.

Once the ground, ionospheric and geomagnetic parameters have been determined, several tests are made on the parameters to determine if the path point should be saved for further processing. A path point is always used if the ground conductivity or the ionospheric profile changes from the values found at the previous point on the path. Additional points are used depending on the ionosphere and the direction and geomagnetic latitude of the point. However, the minimum separation between path points is 100 km. The decisions to use any point depends in large part on how the waveguide mode solutions vary under differing ionospheric conditions and from one geomagnetic condition to another. For example, propagation anisotropy is considerably diminished under daytime conditions as compared to nighttime conditions. Therefore, the variation of the geomagnetic parameters between points on the path is allowed to be larger in daytime than at night. An additional consideration is the balancing of the completeness of the mode searching routine used in "MF" with the speed of the extrapolation routine used in "SW".

BASIC INPUT TO THE SETUP PROGRAM

Basic input to the setup program consists of a case identification, file identification, the transmitter location and frequency, the date and time. The propagation paths may be specified in either of two ways, the maximum distance and a series of bearing angles of the propagation paths or the limits of an area for coverage analysis. A specific propagation path can be defined by using the coordinates of a receiver or by using a single value of bearing. Alternatively, a fan of bearing angles may be specified. The purpose of such a fan is to generate data for paths which will cover a specific operating area or a large geographic area. The simplest fan consists of propagation paths at bearings of 0 through 350°, in increments of 10°. A more complete fan of bearing angles is obtained when the boundaries of an operating area are specified. In this case, the bearing angles are generated in 10° increments sufficient to cover the operating area. Additional bearing angles (with 3° resolution) are generated as necessary in order to ensure that enough paths are generated to include areas of the earth's surface which have ground conductivity less than 10^{-3} S/m. These additional paths are generated on the basis of the ground conductivity map included in the model.

COMMAND FILE GENERATION

The result of the path segmentation described above is a list of points which indicate changes in the ground conductivity or ionospheric profile. The basic strategy of the automated procedure is to rearrange this list so as to produce a set of points which are uniform in conductivity and ionospheric profile. The resulting sequence of segments may not be contiguous with respect to distance. However, the mode extrapolation procedure of "SW" can be employed to quickly obtain the waveguide mode solutions for the segments which have a common ground conductivity and ionosphere. Since the rearrangement of the segments is based on a predetermined order, it is even possible that the first segment to be processed is not at the transmitter. The use of a mode conversion model to obtain the signal strength as a function of distance allows for the fact that adjacent segments along a path may contain different waveguide mode parameters.

The VMS driver is used to control the order of execution of "MF" and "SW". The driver starts at the first point in the list of segments. It sets up an "MF" run using

the appropriate waveguide parameters for the point in question. "MF" is executed and its output is used to provide starting solutions to "SW" which is then executed for the remaining segments which have the same ground conductivity and ionosphere. If "SW" fails to duplicate the input solutions obtained from the "MF" run, then the VMS driver skips the segment. If "SW" fails at some other segment, then the VMS driver starts a new "MF" execution at the point of failure in order to get new starting solutions for "SW". When the end of all of the segments is reached, the output from all of the "SW" executions is rearranged to be in order of increasing distance. This output is then checked for missing segments. If none of the segments have been skipped, then another VMS procedure is executed to delete the command and log files. If any segments have been skipped, then the command and log files are saved for examination and subsequent correction. In any case, upon completion of the current path, the command file for the next path in the set is started. If the current path is the last one, then the mode summing program is executed.

ATMOSPHERIC NOISE

In cases for which a series of propagtion paths are produced, the output from the mode summing program is signal strength along the same series of propagation paths. These data can be used to map the signal strength onto a geographical display for analysis of coverage. A critical component of such analysis is determination of the levels of atmospheric noise within the area of interest. Three models of atmospheric noise are widely used. Two of these are based on the models and data of the CCIR. The first noise model is designated ITS. It is derived from mapping in Universal Time by Zacharisen and Jones (1970). This model is based on CCIR-322 (1963). A new noise model, designated NTIA, has been developed by Spaulding and Washburn (1985). These two models use sets of numerical coefficients to produce values of atmospheric noise at a specified location, season and time. The models are valid from 10 kHz through 1 MHz. The third model of atmospheric noise was developed by Maxwell et al. (1970). This model is designated WGL and is valid only from 10 to 30 kHz. The WGL model uses a data base consisting of thunderstorm activity and a simplified propagation model to compute the strength of atmospheric noise. Even though the propagation model is simplified, the combination of large numbers of thunderstorms and propagation calculations makes determination of atmospheric noise with this model quite slow.

In the sample case to be presented below, the area of interest is the Mediterranean Sea. Maps of noise in dB above 1uv/m for each of the noise models are shown in Figure 3. These maps show contours of constant noise level. The maps for NTIA and WGL are similar but the ITS model is 3 dB higher in the eastern part of the area. It should be noted that this agreement in the Mediterranean Sea is not the case in general. There are regions of the world where the difference among the models is tens of dB.

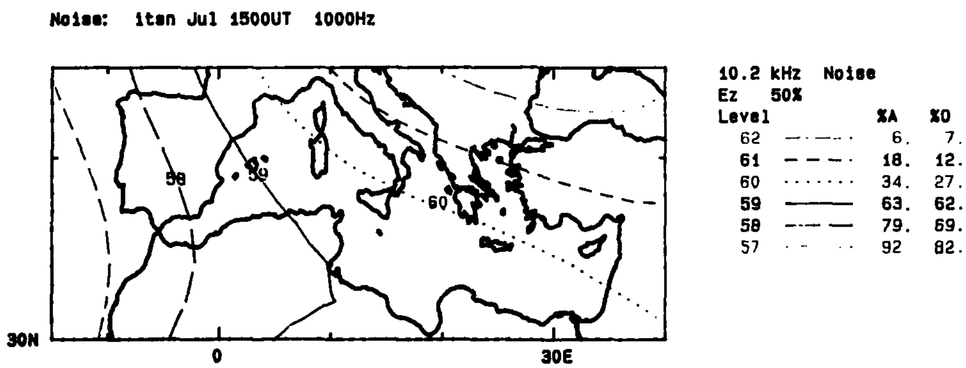
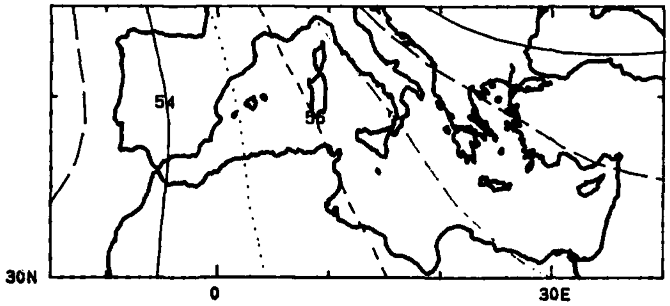


Figure 3a: Map showing atmospheric noise from the ITS model.

Noise: ntia Jul 1500UT 1000Hz

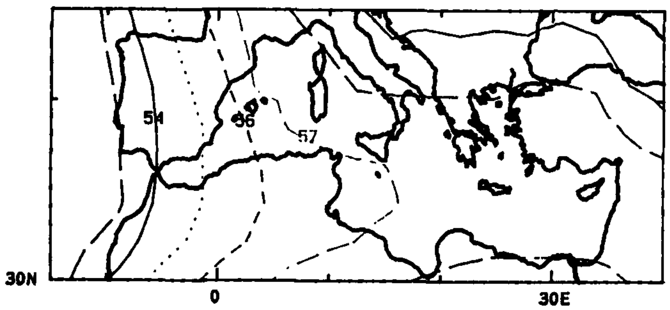


10.2 kHz Noise
Ez 50%

Level		%A	%O
59	————	7.	8.
58	—— ——	19.	10.
57	· · - · · ·	39.	45.
56	— — — ·	55.	56.
55	· · · · · · ·	68.	63.
54	————	80.	71.
53	—— ——	96.	90.

Figure 3b: Map showing atmospheric noise from the NTIA model.

Noise: deco Jul 15. 1988 1500UT 1000Hz



10.2 kHz Noise
Ez 50%

Level		%A	%O
59	————	7.	5.
58	—— ——	24.	15.
57	— · — · ·	59.	57.
56	— — — ·	69.	63.
55	· · · · · · ·	77.	68.
54	————	84.	71.
53	—— ——	88.	75.

Figure 3c: Map showing atmospheric noise from the WGL model.

SAMPLE CASE

A sample case of coverage of the Mediterranean Sea at 10.2 kHz from the OMEGA transmitter located in North Dakota, USA, is presented to illustrate the coverage model. The date and time is chosen to give all day-time conditions between the transmitter and the Mediterranean Sea. Figure 4 shows the Mediterranean Sea and the transmitter. The operating area is shown in the lower right hand corner of the large area. The propagation paths which are used to span the operating area are shown in this figure. The dots which appear along each path indicate the boundaries of individual segments. As discussed above, these boundaries are chosen on the basis of ionospheric, geomagnetic and ground conductivity changes. In particular, for the problem being considered here, the segmentation is principally controlled by changes in ground conductivity. It is noted that Greenland has very low conductivity which produces dramatic effects in the coverage maps to be shown later.

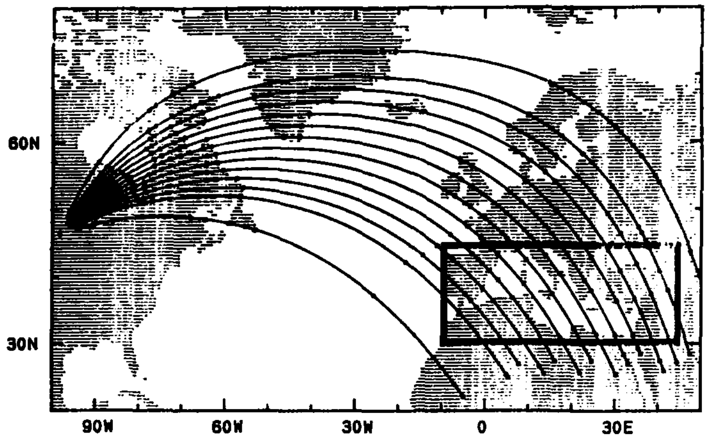


Figure 4: Map showing transmitter and operating area.

Signal strength in dB above 1 uv/m as a function of distance along the most northerly path is shown in the top panel of Figure 5. The corresponding result for the most southerly path is shown in the bottom panel. For reference, in each of these panels are lines indicating the variation of ground conductivity along the path (the line nearest the bottom) and the ionospheric height (the second line from the bottom). In the top panel, the effect of the very low conductivity of Greenland is seen in the rapid drop in signal strength between 3800 and 4300 km.

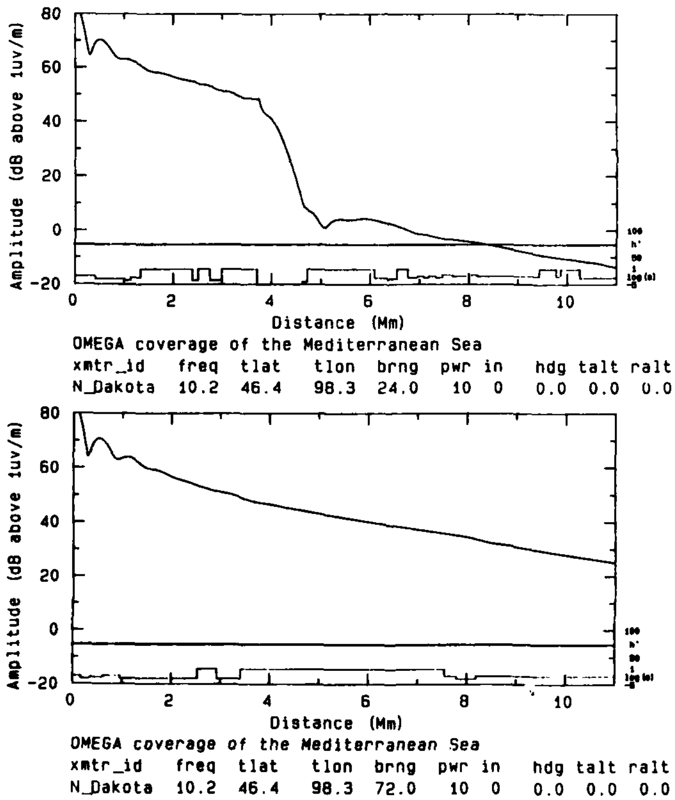
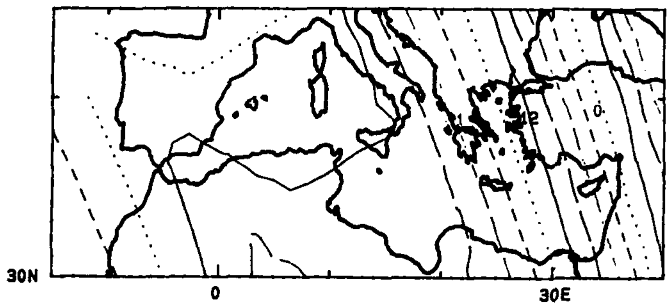


Figure 5: Plot of signal strength vs. distance for two paths.

The effect of the low conductivity of Greenland is even more dramatic in Figure 6 which shows contours of constant signal strength over the operating area. The very close spacing of nearly parallel contours in the eastern half of the area is due to the large signal losses caused by Greenland. These closely spaced contour lines disappear when the paths from the transmitter to the operating area pass below Greenland. Incorporation of the noise data gives maps of signal to noise ratio. The map shown in Figure 7 was generated using the NTIA model. The time availibility is determined by assuming Gaussian statistics.

xmtr_id	lat	lon	dB in	hdg	ht	hr	stn	prfl_id	file
N_Dakota	46.4	98.3	10	00	000	0	0 0.0	All day	ndakota_Mediterr



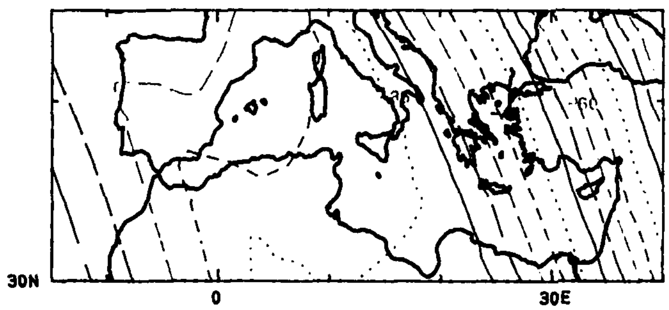
10.2 kHz Signal
Ez 90%

Level	%A	%O
33	4.	9.
30	17.	26.
27	39.	48.
24	60.	67.
21	66.	74.
18	71.	77.
15	74.	81.
12	78.	86.
9	81.	89.

Figure 6: Plot of signal strength over the operating area.

xmtr_id	lat	lon	dB in	hdg	ht	hr	stn	prfl_id	file
N_Dakota	46.4	98.3	10	00	000	0	0 0.0	All day	ndakota_Mediterr

Noise: ntia Jul 1500UT 1000Hz



10.2 kHz S/N
Ez 90%

Level	%A	%O
-24	3.	6.
-27	9.	17.
30	21.	31
-33	34.	42.
-36	52.	59.
-39	64.	69.
-42	66.	74.
-45	71.	78
-48	74.	81.

Figure 7: Plot of signal to noise ratio over the operating area using NTIA.

REFERENCES

Budden, K. G., The Waveguide Mode Theory of Wave Propagation, Logos Press, London, 1961.

CCIR, World Distribution and Characteristics of Atmospheric Radio Noise, Rept. 322, Documents of the Xth Plenary Assembly, Geneva, 1963.

Ferguson, J. A., Ionospheric Profiles for Prediction Nighttime VLF/LF Propagation, NAVOCEANSYSCEN TR 530, 1980.

Ferguson, J. A., F. P. Snyder, Approximate VLF/LF Waveguide Mode Conversion Model, NAVOCEANSYSCEN TD 400, 1980.

Ferguson, J. A., F. P. Snyder, The Segmented Waveguide Program for Longwavelength Propagation Calculations, NAVOCEANSYSCEN TD 1071, 1986.

Maxwell, E. L., D. L. Stone, R. D. Croghan, L. Bell, A. D. Watt, Development of a VLF Atmospheric Noise Prediction Model, Westinghouse Georesearch Laboratory Rept. 70-1H1-VLF NO-R1, 1970.

Morfitt, D. G., Effective Electron Density Distributions Describing VLF/LF Propagation Data, NAVOCEANSYSCEN TR 141, 1977.

Morfitt, D. G., J. A. Ferguson, F. P. Snyder, Numerical Modeling of the Propagation Medium at ELF/VLF/LF, AGARD CP 305, 1981

Morgan, R. R., World-wide VLF Effective Conductivity Map, Westinghouse Electric Corp. Rept. 8013F-1, 1968.

Pappert, R. A., E. E. Gossard, I. J. Rothmuller, A numerical investigation of classical approximations used in VLF propagation, Radio Science 2, 387-400, 1967.

Pappert, R. A., L. R. Shockey, Mode Conversion Progrm for an Inhomogeneous Anisotropic Ionosphere, NELC Interim Report 722, 1972.

Pappert, R. A., J. A. Ferguson, VLF/LF mode conversion model calculations for air to air transmissions in the earth-ionosphere waveguide, Radio Science 21, 551-558, 1986.

Shellman, C. H., A New Version of MODESRCH Using Interpolated Values of the Magnetoionic Reflection Coefficients, NAVOCEANSYSCEN TR 1473, 1986.

Spaulding, A. D., J. S. Washburn, Atmospheric Radio Noise: Worldwide Levels and Other Characteristics, U. S. Dept. of Commerce NTIA Rept. 85-173, 1985.

Zacharisen, D. H., W. B. Jones World Maps of Atmospheric Radio Noise in Universal Time by Numerical Mapping, U. S. Dept. of Commerce, Office of Telecommunications Rept. OT/ITSRR 2, 1970.

A Communications System with an Adaptive Multi-Functional Architecture.

A.H.Kemp.

Hull-Warwick Communications Research Group,
University of Hull,
Hull HU6 7RX.
U.K.

Abstract.

Within NATO's area of operations, high frequency (HF) channels exhibit an inherently variable capacity. Furthermore there is also a wide geographical variety of noise sources; in North America atmospheric noise is significant whilst, in Western Europe, a dominant noise source is co-channel interference. A communications system which can operate across this whole region needs to be both highly adaptive and have a real-time channel evaluation (RTCE) capability. To allow a system to manage multiple adaptive components in an optimum manner whilst the system configuration and RTCE data vary, a multi-functional regime is proposed. This paper describes the design and implementation of such an adaptive, multi-functional communications system.

1. Introduction.

A radio communication system's parameters will tend to vary to some degree with time. The parameters may vary on a short-term basis because of path conditions or on a longer-term basis due to component drift etc. In addition to this the architecture of a system in terms of the level of privacy, degree of error control, and even the required number of users, are factors which can vary with time. The aim of the research programme described in this paper is to design and test a system which can adapt to both short-term and longer-term variability of a communication channel.

In order to produce the type of system indicated above, some form of unifying design concept was deemed necessary. The particular nexus used here is multi-level complementary sequence sets. The synthesis of such multi-level sequences and uncorrelated multi-level sequence sets constitutes novel work and is described in section 2. These sets are integer multi-level codes, unlike Sivaswamy's complex multiphase codes described in reference 1.

Since these complementary sequences are pseudorandom in nature, they can also be used to generate real-time, channel state data,ie for embedded real-time channel evaluation (RTCE). This can then be used to provide the necessary assessment of path quality to initiate control strategies in an adaptive architecture [2].

The uncorrelated nature of the sets allows multi-user communications through the use of code-division multiplexing (CDM). Varying the length of the sequences used will alter the throughput and can be used to give a varying level of error control by introducing increased levels of redundancy. With the possible inclusion of encryption, interleaving, and by varying sequence length, data security can also be adjusted. In addition to these system parameters, multi-level sequence sets allow the degree of diversity to be adaptively varied through altering the number of levels in a particular set. If the levels are coded as different frequencies this will also alter the modulation. The system can be run either synchronously or asynchronously as the required throughput and channel conditions vary. Further, the length of the pseudorandom sequences will change the quality of embedded RTCE data that can be extracted.

The adaptive functions of the system are therefore:

1) Number of users,
2) Privacy/security,
3) Error control,
4) Diversity processing,
5) Modulation,
6) Synchronisation,
7) Quality of RTCE data.

The use of uncorrelated, multi-level complementary sequences can potentially provide all the above attributes to some degree. Hence this forms the basis of a multi-functional system, as defined in reference 3.

The system described here is designed particularly for the HF band, since this is an area where considerable variability of channel conditions occurs and consequently a highly adaptive system is ideally required [4]. It is planned that a meteor-burst orientated version will also be implemented since this is another area suited to a highly adaptive system [5].

2. The Synthesis of Multi-level Complementary Sequences.

2.1 Introduction & Modified Definition.

In reference 6, Golay defines a pair of binary bipolar complementary sequences as two, equal-length sequences which have the property that the number of pairs of like elements with any given separation in one sequence is exactly equal to the number of

pairs of unlike elements with the same separation in the other sequence. This property leads to the characteristic form of the summed non-periodic autocorrelation function (acf) for a pair of binary complementary sequences, viz. that the sum of the individual, non-periodic acf's of the two sequences at corresponding time shifts is always zero, except at the zero-shift (in-phase) position where it takes the value 2N, N being the number of bits in each sequence.

Normally,

$$N = 2^n \quad \text{(n integer)} \tag{1}$$

When the number of sequences in a complementary set is extended beyond two, the original definition is no longer appropriate and a definition directly related to acf properties has been adopted, i.e. that the sum of the individual non-periodic acf's of all members of the set is zero at corresponding shifts except at the zero-shift position where the summed acf will take a value equal to the sum of the squares of all digits in all sequences of the set. This definition is also applicable to sets of non-binary (multi-level) complementary sequences.

For completeness: a pair of uncorrelated complementary sequence sets is defined as a pair of sequence sets for which the non-periodic cross-correlation function (ccf) values for corresponding sequences in each set sums to zero at all corresponding time shifts. This is the property of these sets which makes them suitable for CDM. For example, the two binary sequence sets, each comprising four sequences

	Set (a)	Set (b)	
sequence 1:	+1,+1,+1,+1	-1,+1,-1,+1	
sequence 2:	+1,+1,-1,-1	-1,+1,+1,-1	
sequence 3:	-1,+1,-1,+1	+1,+1,+1,+1	(2)
sequence 4:	-1,+1,+1,-1	+1,+1,-1,-1	

are both complementary and uncorrelated in the sense indicated above. This is demonstrated below by showing their respective acfs and the ccfs of corresponding rows:
The autocorrelation functions of set(a);

row 1:	1,	2,	3,	4,	3,	2,	1	
ow 2:	-1,	-2,	1,	4,	1,	-2,	-1	(3)
row 3:	-1,	2,	-3,	4,	-3,	2,	-1	
row 4:	1,	-2,	-1,	4,	-1	-2,	1	
sum:	0,	0,	0,	16,	0,	0.	0.	

The ccfs between corresponding rows of sets (a) and (b):

row 1:	1,	0,	1,	0,	-1,	0,	-1	
row 2:	-1,	0,	3,	0,	-3,	0,	1	(4)
row 3:	-1,	0,	-1,	0,	1,	0,	1	
row 4:	1,	0,	-3,	0,	3,	0,	-1	
sum:	0,	0,	0,	0,	0,	0,	0.	

2.2 <u>Synthesis Procedures for Binary Sets.</u>

In reference 7, recursive procedures for synthesising pairs of binary complementary sequences are described. Assuming that the length of each sequence in a pair is N, if **A** and **B** are sub-blocks of length N/2, then a pair of complementary sequences can be synthesised from the structures:

(a) +A +B	**(b)** +A -B	**(c)** +A +B	**(d)** -A +B	(5)
+B -A	+B +A	-B +A	+B +A	

For example, if A=+1 and B=+1 then applying 5(a) and 5(b) yields the 2-bit binary sequence sets:

(a) +1 +1	**(b)** +1 -1	(6)
+1 -1	+1 +1	

which are both complementary. Also, the summed non-periodic ccf between corresponding sequences of 6(a) and 6(b) is zero for all time shifts, thus demonstrating that the two sets are uncorrelated. Note that the structures 5(c) and (d) also provide a pair of uncorrelated complementary sets, whereas (a)-(c), (a)-(d), (b)-(c), (b)-(d) do not.

If A and B are now redefined as 6(a) and 6(b) respectively, and the structure of 5(a) and 5(b) is employed again, two uncorrelated complementary sets, each comprising four 4-bit sequences, result.

i.e.

	+1 +1 +1 -1		+1 +1 -1 +1	
(a)	+1 -1 +1 +1	(b)	+1 -1 -1 -1	(7)
	+1 -1 -1 -1		+1 -1 +1 +1	
	+1 +1 -1 +1		+1 +1 +1 -1	

The procedure can again be applied recursively with 7(a) and 7(b) now representing the basic seed blocks A and B.

2.3 **Synthesis Procedures for Multi-level Sets.**

The same basic recursive procedure can be employed to synthesise multi-level (non-binary) complementary sets. If the integers 1 and 2 are chosen as A and B respectively in the structure 5(a), and the integers 3 and 4 as A and B respectively in the structure of 5(b), the sets:

$$\text{(a)}\quad \begin{matrix} 1 & 2 \\ 2 & -1 \end{matrix} \quad \text{and} \quad \text{(b)}\quad \begin{matrix} 3 & -4 \\ 4 & 3 \end{matrix} \qquad (8)$$

are produced each of which are complementary. However, they are not uncorrelated because of the different digit weightings in the 2 sets. If 8(a) and 8(b) are now redefined as A and B in the structures of 5(a) and 5(b), the following sequence sets result:

$$\text{(a)}\quad \begin{matrix} +1 & +2 & +3 & -4 \\ +2 & -1 & +4 & +3 \\ +3 & -4 & -1 & -2 \\ +4 & +3 & -2 & +1 \end{matrix} \qquad \text{(b)}\quad \begin{matrix} +1 & +2 & -3 & +4 \\ +2 & -1 & -4 & -3 \\ +3 & -4 & +1 & +2 \\ +4 & +3 & +2 & -1 \end{matrix} \qquad (9)$$

These are both complementary and uncorrelated; the latter property arises because the sets now have the same distribution of element weights. Again, the synthesis procedure can be applied recursively to give sets of larger sequences.

2.4 **Synthesis of Odd-Length Multi-level Sequence Sets.**

If columns of "all-zeros" are added to sets 9(a) and 9(b) as shown below:

$$\text{(a)}\quad \begin{matrix} +1 & +2 & +3 & -4 & 0 \\ +2 & -1 & +4 & +3 & 0 \\ +3 & -4 & -1 & -2 & 0 \\ +4 & +3 & -2 & +1 & 0 \end{matrix} \qquad \text{(b)}\quad \begin{matrix} 0 & +1 & +2 & -3 & +4 \\ 0 & +2 & -1 & -4 & -3 \\ 0 & +3 & -4 & +1 & +2 \\ 0 & +4 & +3 & +2 & -1 \end{matrix} \qquad (10)$$

then these two sets are still complementary and their uncorrelated properties are also maintained. If a term-by-term addition is then performed on 10(a) and 10(b) a further set comprising of four 5-digit sequences results, ie.

$$\begin{matrix} +1 & +3 & +5 & -7 & +4 \\ +2 & +1 & +3 & -1 & -3 \\ +3 & -1 & -5 & -1 & +2 \\ +4 & +7 & +1 & +3 & -1 \end{matrix} \qquad (11)$$

This set is also complementary. By augmenting different sets of complementary sequences with all-zero columns, any length of sequences odd or even, can be generated by term-by-term addition.

Generally, term-by-term addition of two or more sets of complementary sequences of the same dimensions will result in a further complementary set of the same dimensions but different level weightings.

2.5 **Synthesis of Uncorrelated Sets of Multi-level Sequences.**

Two sets of uncorrelated complementary sequences can be generated by simply using structures 5(a) and 5(b) or structures 5(c) and 5(d) as previously shown. However this procedure can only generate two uncorrelated sets and for the multi-user system being considered, more sets are required. The problem is to introduce the other structures into the construction of the sets. This is done by first using two of the structures to generate seed sets, which in turn are then employed by the other two structures to generate the uncorrelated, multi-level, complementary sequence sets. For instance, eight uncorrelated, 8-level, complementary sequence sets may be generated as follows:

(i) Using digits 1 and 2 form two, 2-by-2 sets, with constructions 5(a) and 5(b). Then recursively use these 2 sets to generate four, 4-by-4 sets, one by each construction 5(a) to 5(d). Label these four, 4-by-4 sets, A1 to A4.
(ii) Using digits 3 and 4 form two, 2-by-2 sets, with constructions 5(c) and 5(d). Then recursively use these 2 sets to generate four, 4-by-4 sets, one by each construction 5(a) to 5(d). Label these four, 4-by-4 sets, B1 to B4.
(iii) Form two, 8-by-8 sets, by constructions 5(a) and 5(b), using A1 and B2 as the seed sets.
Form two, 8-by-8 sets, by constructions 5(a) and 5(b), using A2 and B1 as the seed sets.
Form two, 8-by-8 sets, by constructions 5(c) and 5(d), using A3 and B4 as the seed sets.
Form two, 8-by-8 sets, by constructions 5(c) and 5(d), using A4 and B3 as the seed sets.

This procedure is illustrated below.

Stage (i).

$$\text{(a)}\quad \begin{matrix} 1 & 2 \\ 2 & -1 \end{matrix} \qquad \text{(b)}\quad \begin{matrix} 1 & -2 \\ 2 & 1 \end{matrix}$$

$$\text{(A1)}\quad \begin{matrix} 1 & 2 & 1 & -2 \\ 2 & -1 & 2 & 1 \\ 1 & -2 & -1 & -2 \\ 2 & 1 & -2 & 1 \end{matrix} \qquad \text{(A2)}\quad \begin{matrix} 1 & 2 & -1 & 2 \\ 2 & -1 & -2 & -1 \\ 1 & -2 & 1 & 2 \\ 2 & 1 & 2 & -1 \end{matrix}$$

(A3)	1	2	1	-2	(A4)	-1	-2	1	-2
	2	-1	2	1		-2	1	2	1
	-1	2	1	2		1	-2	1	2
	-2	-1	2	-1		2	1	2	-1

Stage (ii).

(c)	3	4	(d)	-3	4
	-4	3		4	3

(B1)	3	4	-3	4	(B2)	3	4	3	-4
	-4	3	4	3		-4	3	-4	-3
	-3	4	-3	-4		-3	4	3	4
	4	3	4	-3		4	3	-4	3

(B3)	3	4	-3	4	(B4)	-3	-4	-3	4
	-4	3	4	3		4	-3	4	3
	3	-4	3	4		-3	4	3	4
	-4	-3	-4	3		4	3	-4	3

Stage (iii).

1	2	1	-2	3	4	3	-4
2	-1	2	1	-4	3	-4	-3
1	-2	-1	-2	-3	4	3	4
2	1	-2	1	4	3	-4	3
3	4	3	-4	-1	-2	-1	2
-4	3	-4	-3	-2	1	-2	-1
-3	4	3	4	-1	2	1	2
4	3	-4	3	-2	-1	2	-1

1	2	1	-2	-3	-4	-3	4
2	-1	2	1	4	-3	4	3
1	-2	-1	-2	3	-4	-3	-4
2	1	-2	1	-4	[illegible]	4	-3
3	4	3	-4	1	[illegible]	1	-2
-4	3	-4	-3	2	[illegible]	2	1
-3	4	3	4	1	[illegible]	-1	-2
4	3	-4	3	2	1	-2	1

1	2	-1	2	[illegible]	4	-3	4
2	-1	-2	-1	-4	3	4	3
1	-2	1	2	-3	4	-3	-4
2	1	2	-1	4	3	4	-3
3	4	-3	4	-1	-2	1	-2
-4	3	4	3	-2	1	2	1
-3	4	-3	-4	-1	2	-1	-2
4	3	4	-3	-2	-1	-2	1

1	2	-1	2	-3	-4	3	-4
2	-1	-2	-1	4	-3	-4	-3
1	-2	1	2	3	-4	3	4
2	1	2	-1	-4	-3	-4	3
3	4	-3	4	1	2	-1	2
-4	3	4	3	2	-1	-2	-1
-3	4	-3	-4	1	-2	1	2
4	3	4	-3	2	1	2	-1

1	2	1	-2	-3	-4	-3	4
2	-1	2	1	4	-3	4	3
-1	2	1	2	-3	4	3	4
-2	-1	2	-1	4	3	-4	3
3	4	3	-4	1	2	1	-2
-4	3	-4	-3	2	-1	2	1
3	-4	-3	-4	-1	2	1	2
-4	-3	4	-3	-2	-1	2	-1

-1	-2	-1	2	-3	-4	-3	4
-2	1	-2	-1	4	-3	4	3
1	-2	-1	-2	-3	4	3	4
2	1	-2	1	4	3	-4	3
-3	-4	-3	4	1	2	1	-2
4	-3	4	3	2	-1	2	1
-3	4	3	4	-1	2	1	2
4	3	-4	3	-2	-1	2	-1

-1	-2	1	-2	3	4	-3	4
-2	1	2	1	-4	3	4	3
1	-2	1	2	3	-4	3	4
2	1	2	-1	-4	-3	-4	3
-3	-4	3	-4	-1	-2	1	-2
4	-3	-4	-3	-2	1	2	1
-3	4	-3	-4	1	-2	1	2
4	3	4	-3	2	1	2	-1

1	2	-1	2	3	4	-3	4
2	-1	-2	-1	-4	3	4	3
-1	2	-1	-2	3	-4	3	4
-2	-1	-2	1	-4	-3	-4	3
3	4	-3	4	-1	-2	1	-2
-4	3	4	3	-2	1	2	1
3	-4	3	4	1	-2	1	2
-4	-3	-4	3	2	1	2	-1

The eight, 8-level, 8-by-8 sets, formed as shown above, are all uncorrelated, complementary sequence sets.

3. Adaptive Functions.

3.1 Number of Users.

Since uncorrelated sets of complementary sequences have been generated, this feature can be used to provide multi-user communications. Two obvious ways of using the non-interfering properties of uncorrelated sets exist: the first is to sum the levels of the sets before modulation; the composite sequences thus formed will still activate the correct matched filters, whether or not the same relative positions in time are maintained [see Fig.1(a)]. The second method is to simply transmit each sequence regardless of the other users and allow the levels to sum, or otherwise, during transmission over the propagation path [see Fig.1(b)]. These procedures both imply an assumption of path linearity.

The first of these two methods requires an increased number of modulation states, bit synchronisation between users, and all users to be connected to a common transmitter station. However, if all these requirements can be met the method provides multi-user communications with no increase in co-channel interference/noise.

Alternatively, the second method, with each user transmitting independently as suggested above, does not have any of the first methods requirements. However, with this second method, more power will be transmitted simultaneously in the same bandwidth. The uncorrelated nature of the sequence sets means that ideally, the increased level of power will not compromise system performance: however in practice some performance degradation may be expected.

3.2 Error Control.

The level of error control required by each user of the system may vary independently with time. By altering the length of the complementary sequences, the amount of redundancy and hence the error control potential will change. Clearly, this is a function which must interact with the RTCE data to ideally maintain a constant output probability of error. This will, of course, result in comparatively high data rates when conditions are favourable, eg.in a meteor burst system when a large ionisation cone occurs.

3.3 Privacy/Security.

The ability of an eavesdropper to usefully intercept the information contained in a transmission depends on his level of knowledge of the transmission parameters. Data that can potentially contribute to successful interception includes frequency, modulation type, baud rate, and the channel/source encoding scheme (including knowledge of any interleaving being employed). This data can normally be extracted by monitoring the transmission for a sufficiently long period if the interceptor has a priori knowledge of expected traffic. Clearly, the longer the monitoring required before useful interception of the transmission can take place, the higher the level of privacy/security.

In a communication system employing complementary sequences, the level of privacy can be improved in a number of ways:

(i) By varying the number of frequency tones used in a given bandwidth, and hence altering the modulation format and the frequency data.

(ii) Interleaving the sequence set values before and after frequency coding, thus providing two layers of interleaving.

(iii) Using complementary sequences which inherently require 'a priori' knowledge for the information extraction. For these, decoding would normally take the form of matched filtering.

(iv) Varying the baud in response to user requirements and channe' 'itions. This implies that a potential interceptor will need to asses system protocols for changes in baud, synchronisation, etc.

3.4 Diversity Processing.

As is discussed in section 4, mapping the multi-level sequences to multiple frequencies will, through in-band frequency spreading, provide a form of frequency diversity which will help protect against frequency selective fading. Varying the number of levels and hence the number of frequencies, will change the order of diversity.

Two levels of interleaving can be achieved by firstly interleaving sequence elements before combination and then again before modulation. This time reordering of the information provides time diversity.

3.5 Modulation.

The number of levels in the complementary sequence sets employed may be varied with time as the required degree of frequency diversity changes. In the multiple frequency-shift-keyed (MFSK) system considered here, this will alter the modulation format used. The number of tones will also be a function of the level of privacy required and the error control scheme in use.

3.6 Synchronisation.

The system designed is set up to work asynchronously taking advantage of the acf peaks which occur in a matched filter demodulation scheme (thereby deriving a form of "modulation derived synchronisation") [8]. One envisaged development however is to modify the system for synchronous operation so that, under very poor channel conditions, this mode can be adaptively selected to allow some throughput to be maintained. A suitable "code assisted bit synchronisation" (CABS) scheme has already been derived [8] and it is hoped to incorporate this in the overall system.

3.7 Quality of RTCE.

The length of sequences used and the number of levels within a sequence both affect the decoded acf peak amplitude at the matched filter output. Consequently when channel conditions degrade and the sequence length is extended to maintain the desired probability of error (error rate), the approximation to the perfect impulse function will become more precise. This means that, as conditions deteriorate, the quality of the available RTCE data will improve automatically. If further information is deemed necessary, then the number of sequence levels employed can also be increased which will further enhance the quality of the RTCE data.

4. Implementation Considerations.

It was felt that the multi-level complementary sequences discussed in section 2, could best be utilised by mapping them to different frequencies in an MFSK modulation scheme. This coding means that their pseudorandom nature will give in-band pseudorandom spectral spreading. This will reduce problems due to in-band frequency selective fading. The multi-user capability will be facilitated by the additional use of CDM or level combination before modulation. This multi-user capability is achieved by detecting the frequencies at the receiver and then matched filtering the resultant waveforms.

For this scheme, eight level sets (ie. +/- 4) have been synthesised and an eight-tone asynchronous modem built which allows any number of the tones to be simultaneously generated and detected.

4.1 Modem Design.

The modem constructed is completely software implemented and runs on an IBM compatible PC with a LSI TMS 320C25 PC board installed in the PC. It is designed to interface with an ICOM IC-735 transceiver. This requires the modem to work within a bandwidth of 2.3 kHz for the baseband signal and a maximum signal amplitude of 500mV (peak. to peak.)[9]. Since one or more of the modulating tones can be present at any one time each tone output is of 62.5mV amplitude (peak. to peak.) to ensure uniform detection across the bandwidth and a peak amplitude for all 8 tones of 500mV. The input is sampled at 6.133 kHz and the output samples generated at this rate. This is over two times the 2.3 kHz bandwidth and so confidently satisfies the Nyquist criterion. The modem has a nominal baud rate of 256 bits/sec., which will finally also be adaptable to 128 and 64 bits/sec.

A matched filter algorithm filters the 24-sample input buffer 8 times, once for each of the 8 tone frequencies. A decision based on the resultant peak height is made and the final tone detection decision is written to an external file and to a screen display. Since any number of tones (0-8) can be present at any one time, the ccf peak detected by the matched filter must be compared with a threshold level within each matched filter detection section. This processing occurs between each sample. In addition to this demodulation processing, the modulation signal must also be synthesised between each sample.

The modulation signal is derived by having a 48 sample series of sine wave values held in memory and using an addressing algorithm to obtain the correct profile. By having the desired address held in a location and incrementing this address by an integer, which is a multiple of the fundamental sine wave contained in memory, the correct frequency output is obtained. For multiple tone output, an address location is used for each tone, and the value at each address location is simply summed prior to output. A block diagram of this modem structure is shown in figure 2.

4.2 Multi-Functional Adaption.

This section of the work has not yet been tested since at present testing is concentrating on confirming the previously postulated attributes of multi-level, complementary, sequence sets. The envisaged regime to provide a multi-functional environment however is clearly a major crux of this research project.

The envisaged implementation of a multi-functional adaptive regime will assign an ascending order of importance to the different adaptive functions, this order of impor-

tance being varied by inter-action of the system manager. The "order of importance figure" (OIF) will then be multiplied by a figure derived from the working system parameters (WSP); eg, if the error control was deemed as the most important attribute it would have OIF of 7. If 20 errors then occurred in the consideration period the WSP would equal 20. The product of these would give 140. This "product figure" (PF) would be assessed so that if it exceeded some predetermined threshold then the system would adapt to reduce it. For example, in the case just given, if the threshold was 100 then since the PF has exceeded this the system must adapt. In particular the level of redundancy, or number of modulation frequencies would be increased until the PF dropped below the threshold. Similarly the PF would be assessed so that if it were below a lower threshold the system would adapt to increase the throughput until the error control measured was within the bounds required. To make the system truly multi-functional however the privacy/security also would have an OIF which would be taken into account when considering whether to increase the redundancy or increase the number of modulation frequencies in order to adjust the error rate, ie. increasing the number of modulation frequencies would increase the degree of privacy as well as reducing the error rate.

The particular interaction of the 7 functions, introduced in section 1, to best provide a multi-functional [3] scheme has yet to be decided. This interaction is clearly very dependent on the results of the tests on multi-level complementary sequence set attributes. However the previous paragraph shows one possible manner in which quantification of the system parameters and useful combination in a multi-functional sense could take place.

5. Results.

Only limited tests have taken place to date since the research is still at an early stage. The construction and validation of a suitable modem capable of simultaneously generating and detecting any number of tones is presently taking place. This modem feature is important to allow exploitation of the properties of the developed sets. However without the modem it has been possible to verify the theoretical properties of multi-level complementary sequence sets. The autocorrelation functions and summed autocorrelation function of set (A1) of section 2.5 is shown in figure 3(a). Figure 3(b) shows the cross-correlation between sets (A1) and (A2) of section 2.5. The two graphs shown in figure 3 illustrate the ideal characteristics of complementary sequence sets for CDM.

By introducing errors into the generated sets, some indication of performance in a noisy environment can be estimated. Two sets have been formed for this. Set 4(i) is the result of removing one of the levels in set (A1) from section 2.5. The level removed was 1 this has been replaced by 0. Set 4(ii) is formed in a similar manner from set (A2) of section 2.5.

The acf of set 4(i) is shown in figure 4(a). It can been seen from figure 4(a) that a considerable detection window still exists between the acf peak and sidelobes, although a quarter of the information has been removed. Figure 4(b) shows the ccf between sets 4(i) and 4(ii) and illustrates the low level of corruption the complete removal of one level creates.

6. Acknowledgement.

The author wishes to thank the British Science and Engineering Research Council (SERC) for their funding of this research programme.

Author.

Name: A.H.Kemp.

Address: Hull-Warwick Communications Group.
Dept. of Electronic Engineering.
University of Hull.
Hull HU6 7RX.
U.K.

7. **References.**

1)**Sivaswamy, R.**:"Multiphase Complementary Codes.":IEEE Transactions on Information Theory, Vol. IT-24, No.5, September 1978.

2)**Darnell, M.**:"Channel evaluation techniques for dispersive communications paths.":NATO ASI 'E',"Communication Systems and Random Process Theory."pp425-60. vol 25. 1978.

3)**Darnell, M. & Honary, B.**:"Multi-functional coding schemes applicable to secure communication." :Proc. IEE Conf. on "Secure Communication Systems."CP-269,pp74-81,London,1986.

4)**Aarons, J.**:"High Frequency Communications :An Introduction.", AGARD, Conference on "Propagation Impact on Modern HF Communications System Design." :Proceedings AGARD Conference no.145, Brussels, 1986.

5)**Watson, K.**:"A Digital Meteor-Burst Communications (MBC) Propagation Path Simulator.", AGARD, "Conference on Propagation Effects and Circuit Performance of Modern Military Radio Systems with Particular Emphasis on those Employing Bandspreading." :Proc. AGARD Conf. no.442, Paris, 1988.

6)**Golay, M.J.E.**:"Complementary Series," :IRE Trans. on Information Theory, Vol IT-7, pp82-87, 1961.

7)**Darnell, M.**:"Principles and Applications of Binary Complementary Sequences", International Symposium on Theory & applications of Walsh & other non-sinusoidal functions, Hatfield Polytechnic, 1975.

8)**Honary B. & Darnell M.**:"New Synchronisation Techniques Applicable to Multi-Tone Transmission Systems.",AGARD, Conf. on "Propagation effects an circuit performance of modern military radio systems with particular emphasis on those employing band spreading." :Proc. AGARD Conf. no.442, Paris, 1988.

9)**ICOM**:IC 735 users reference manual.

Figures.

Figure 1.

(a) (b)

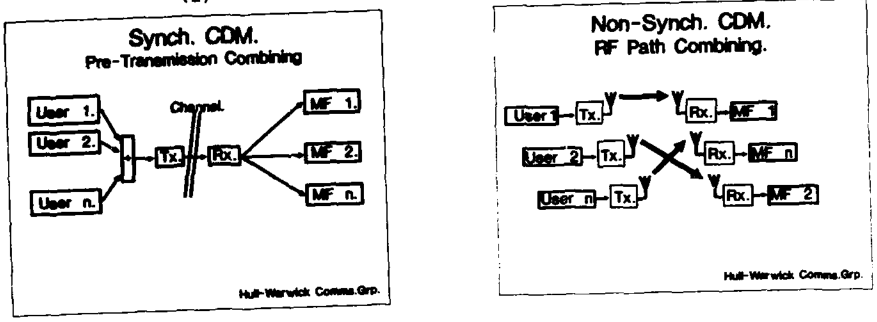


Figure 2.

Block Diagram of Modem.

Demodulation Section.

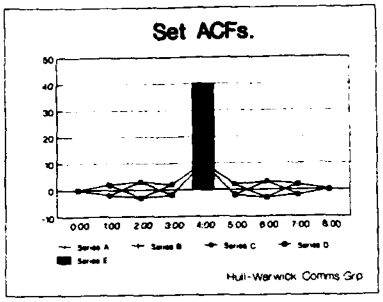
```
                                          Matched    Output
                                          filters.   detection.

                                           ---       ---
                                        --| 1 |--| 1 |--
                                       |   ---       ---   |
                                       |--| 2 |--| 2 |--
                                       |   ---       ---   |
                                       |--| 3 |--| 3 |--
                                       |   ---       ---   |   --
 |Demodulation|__|24 sample      |_|--| 4 |--| 4 |--__|PC|
 |   input.   |  |input buffer.|  |   ---       ---   |   --
  ------------    -------------    |--| 5 |--| 5 |--
                        |          |   ---       ---   |
                        |          |--| 6 |--| 6 |--
                                   |   ---       ---   |
                 |6.133kHz|        |--| 7 |--| 7 |--
                 | clock. |        |   ---       ---   |
                  --------          --| 8 |--| 8 |--
                                       ---       ---
```

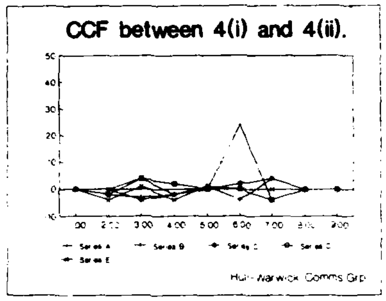
Modulation Section.

```
 ---------------------------------
| Required tones specified        |
|  from an external file.         |
 ---------------------------------
                |
 ---------------------------------
|     Increment values            |
|         selected.               |
 ---------------------------------
                |
 ---------------------------------
|     Value addresses             |
|         calculated.             |
 ---------------------------------
                |
 ---------------------------------
|   Values added to the           |
|      accumulator.               |
 ---------------------------------
                |
 ---------------------------------
|    Total  modulation            |
| signal output from the          |
|    accumulator.                 |
 ---------------------------------
```

Figure 3. **Summed ACF of set (A1) and summed CCF between sets (A1) and (A2).**

(a) 

Series A-D are acfs of sequences 1-4.
Series E is the summed acfs.

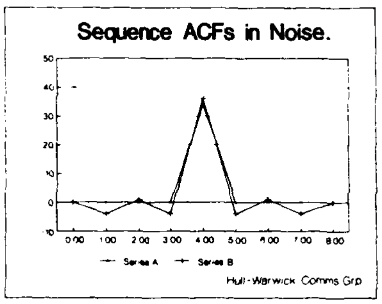
(b) 

Series A-D are individual sequences ccfs.
Series E is the summed ccfs.

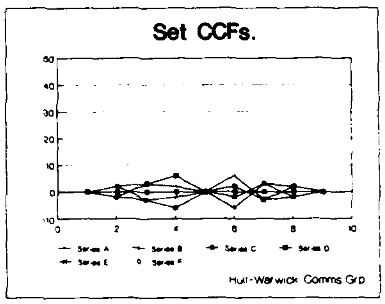
Figure 4.

Sets 4(i)				and 4(ii):			
0	2	0	-2	0	2	-1	2
2	-1	2	0	2	-1	2	-1
0	-2	-1	-2	0	-2	0	2
2	0	-2	0	2	0	2	-1

ACF of set 4(i) and ccf between sets 4(i) and 4(ii).

(a) 

Set 4(i).
Series A : acf of set 4(i).
Series B : acf of set 4(ii).

(b) 

Sets 4(i) and 4(ii).
Series A-D are individual sequences ccfs.
Series E is the summed ccfs.

DISCUSSION

C. GOUTELARD, FR
English Translation

The complementary sequences proposed in your system have been known for a long time and have not been much used, except for radar techniques. They have two main disadvantages for telecommunications. First, separation of the two sequences requires a break in the transmission which imposes technological constraints and reduces the data rate. *Second*, *cross*-correlation properties are not particularly good compared to other sequences or codes - i.e. Gold's codes - or others which do not feature the above disadvantages. Did you carry out a comparative study showing the advantages of your system, and can you situate your sequences with respect to the usual asymptotic limits taken as a reference?

AUTHOR'S REPLY

Golay derived original binary complementary sequences; reported in 1961. The work I have done derives the more general case of multi-level complementary sequence sets and uncorrelated sets of these. These multi-level complementary sequences are novel and have not previously been used. The two drawbacks you mention are invalid. Firstly, each of the component sequences in the set can be used to transmit data. Also each sequence is pseudo-random and hence provides impulse response data of a system if required. Secondly, the summed cross-correlation properties between uncorrelated multi-level complementary sets are perfect, i.e. zero for all time shifts. This provides ideal code division multiplexing characteristics.

AIRLAND BATTLEFIELD ENVIRONMENT (ALBE) DEMONSTRATION PROGRAM

James E. Allen
U.S. Army Engineer Topographic Laboratories
Fort Belvoir, Virginia 22060-5546
USA

ABSTRACT

The U.S. Army recognizes WEATHER and TERRAIN as the two most important factors affecting the outcome on the modern battlefield. The AirLand Battlefield Environment (ALBE) Demonstration Program, a joint effort involving the U.S. Army Corps of Engineers and the U.S. Army Materiel Command, is developing and demonstrating the software, sensors and methodology necessary to integrate a-priori and near-real-time environmental data into map products. These map products, or Tactical Decision Aids (TDAs), will provide the battlefield commander an increased capability to make tactical decisions and thereby defeat the enemy.

INTRODUCTION

Battlefield commanders have only limited access to environmental information during the battle planning process and virtually no access to such information during the battle. To effectively fight an enemy which may be superior in number, move rapidly in all types of weather, use smoke for concealment and use nuclear, biological or chemical weapons, a commander needs to consider all pertinent factors, including environmental effects, to make the most effective tactical decisions. The manual methods used to acquire and analyze environmental data are both time and labor intensive and do not provide the commander with the information he needs on environmental effects.

ALBE TDAs show commanders and their staffs the effects of terrain and environmental factors on both friendly and threat equipment, weapon systems and operations. TDAs do not make decisions themselves, rather they supplement the tactician's knowledge base and help him during the decision making process. TDAs are generated from the synergism of a-priori digital elevation data, weapons systems performance data, force equipment data, structures data and near-real-time environmental sensor data. TDAs provide an invaluable tool in the formulation and execution of both pre-battle and during-battle tactical decisions.

With the increase in the capabilities and the decrease in both the size and the cost of computer systems, the exploitation of the computer on the battlefield is here. Planned development and fielding of systems such as the Maneuver Control System (MCS), the All Source Analysis System (ASAS) and the Digital Topographic Support System (DTSS) will provide the battlefield commander the capability to acquire, analyze and disseminate intelligence and environmental data in a timely manner.

To best exploit the capabilities of new computer systems for maximum tactical advantage, the U.S. Army Corps of Engineers initiated the AirLand Battlefield Environment (ALBE) program. This program is a joint effort involving the Waterways Experiment Station (WES), the Cold Regions Research and Engineering Laboratory (CRREL), the Engineer Topographic Laboratories (ETL) and the Atmospheric Sciences Laboratory (ASL) of the U.S. Army Materiel Command (AMC).

TEST-BED

The ALBE test-bed is composed of two central processing units (CPUs). Each CPU is housed in separate Integrated Command Post shelters which are in turn mounted on the back of Commercial Utility Cargo Vehicles (CUCVs). The test-bed hardware is loaded inside the shelter during transport and can be removed and setup in a tent for use in two hours.

System Hardware

Both the weather and the terrain portions of the ALBE test-bed are built around ruggedized MicroVax II CPUs. The weather CPU is used primarily to collect, store and process surface and atmospheric sensor data, to store and process historical climatological data and to generate weather-intensive TDAs. The terrain CPU is used to store and process elevation, force equipment and structural information data, to create, store and process terrain feature data, and to generate terrain-intensive TDAs. An Ethernet connecting the two CPUs allows for the transfer of data, software and products.

Both CPUs have nearly identical data storage devices. Each has three 380 Mb hard disk drives, one 95 Mb cartridge tape and one 44 Mb 9-track magnetic tape. The weather CPU has one and the terrain CPU has two 200 Mb 5 1/4" optical disks.

The terrain CPU has three and the weather CPU has two medium- resolution (640 X 480) graphics terminals. Both CPUs have one high-resolution (1024 X 1280) graphics processor. The medium-resolution terminals are the primary work stations and the high-resolution graphics processor is used to generate softcopy TDA products. Output devices for graphics and text are identical on both CPUs. Each has a Seiko Thermal Transfer

color copier, an HP 7580B eight pen plotter (large format) and an HP 2934A Dot Matrix printer for text reports.

In addition, the terrain CPU is equipped with a digitizing table for digitizing hardcopy Tactical Terrain Analysis Data Base (TTADB) products provided by the Defense Mapping Agency (DMA) and a high resolution Charge Coupled Device (CCD) camera for generating background maps for the TDAs.

System Software

The operating system for the ALBE Test-bed is MicroVMS, the command language is Digital Command Language (DCL), the assembly language is MACRO and the file editing software is VAX EDT text editor. All four are compatible with their counterparts found on the Digital main-frames such as the VAX 785.

TDA software is written in several languages; however, most is written in VAX FORTRAN 77. Other languages used are VAX C and VAX PASCAL.

Graphics software for generating the products is written in GKS PVI GK - 2000.

A new Geographic Information System (GIS) is being developed for the demonstration program that will provide the capability to quickly and efficiently manipulate large data bases such as Defense Mapping Agency (DMA) Tactical Terrain Data (TTD), to concurrently process and display raster and vector data, to perform batch processing of large jobs in background and to directly access the GIS software code.

ENVIRONMENTAL SENSORS

Four different sensors are used to provide soil and weather information to the ALBE software. During an exercise, three of these sensors are located on the ground in those areas which provide representative weather information and which maintain a radio line of sight with the test-bed. Information from each of these three sensors is obtained at 15 minute intervals thereby ensuring the TDA software uses the most recent weather and soil information. The fourth sensor is balloon mounted to provide upper-air weather data.

A Soil Moisture Sensor (SMS) provides both soil moisture and soil temperature data and is critical to the Mobility and Route Planning TDAs. The Present Weather System (PWS) and the Meteorological Sensor Package (MSP) (developed by ASL) provide valuable low level weather information in support of all TDAs. PWS provides data on visibility, precipitation rate and type, and fog/haze. MSP provides data on wind speed and direction, temperature, pressure, illumination, accumulated precipitation, humidity, and soil temperature.

TACTICAL DECISION AID SOFTWARE

At present, ten TDAs have been defined for demonstration in the ALBE program. All of these TDAs account for both terrain and weather conditions and are described below.

Flight Hazards products identify those areas in the atmosphere where aircraft and personnel are subjected to extreme weather conditions; to nuclear, biological and chemical agents; and to direct fire. Aerial Detection products identify those positions where aircraft can first detect and recognize air and ground targets and can perform non-line-of-sight (NLOS) operations.

Mine Use products identify optimum locations for the emplacement of mines and provide a prediction of the mine effectiveness. The Mine Location predicts the probable location of enemy mines and provides an estimate of the time required to breach or circumnavigate the mines. Obstacles products identify the actual or possible location and effectiveness of natural and man-made obstacles.

Mobility products provide cross-country movement, on-road movement and gap crossing (fords and tactical bridging) predictions for threat and U.S. vehicles. Additionally, an interactive Route Planner can be used with the Mobility products to determine route traversing time and distance.

NBC products predict the spread of NBC contaminants, generates warning messages of NBC conditions and provide information on the effects of Mission Oriented Protective Posture (MOPP) gear on soldier tasks. Smoke products provide predictions of the type, location and amount of smoke generating capability required to perform a specified screening mission.

EO Sensor products provide predictions on the effectiveness of various threat and U.S. EO devices such as optical, infrared (near and thermal), and image intensifiers.

FIELD EXERCISES

A series of exercises have been initiated in which TDAs will be tested and evaluated as to their accuracy, responsiveness and applicability to the needs of commanders and their staffs; and as to the speed and ease of their production versus present day manual methods. The first exercise took place in December 1987 at Ft. Lewis, Washington, during the Command Post Exercise (CPX) Cascade Peake IV. Over 160 products were

successfully generated during this exercise. The second and third exercises will take place in spring and fall 1990. The spring exercise will be a Division level exercise and the fall exercise will be conducted at Corps level. Plans for both of these exercises are to provide input early enough to effect planning operations.

PROTOTYPE INTEGRATED METEOROLOGICAL SYSTEM (IMETS)

The prototype IMETS will be integrated into the ALBE program to evaluate the benefits of wide area weather coverage for both present and forecast conditions. According to the INTEGRATED METEOROLOGICAL SYSTEM (IMETS), PROOF OF CONCEPT PLAN - DRAFT, prepared by ASL, 1 FEB 89, IMETS, when fielded will:

(1) Automate the collection of weather data on the battlefield

(2) Improve the timeliness and accuracy of battlefield weather forecasts and weather warnings

(3) Depict weather effects on both friendly and enemy personnel, weapons, equipment and operations in a form easily understood by tactical commanders during combat

Data provided by the prototype IMETS will include weather data and weather maps from Air Force Global Weather Central (AFGWC); surface data from the MSP, PWS and SMS sensors, and upper air data from either the LAMS or Vaisala Upper Air system or both.

AFGWC weather information includes forecasts for precipitation and cloud cover; maps depicting surface pressure, fronts, cloud cover, precipitation and temperature extremes; hourly information on air temperature, dew point, winds, pressure, visibility, precipitation intensity (light-moderate-heavy-very heavy) and accumulated precipitation. Additionally, weather reports are updated from reports and forecast weather warning advisories are provided.

The addition of data from AFGWC will provide the ALBE software access to several data sources based upon mission need. For long term planning, TDAs based upon climatological data can be generated. For planning operations in the near future, for wide area operations or for deep strike operations; AFGWC data and upper air data can be used to generate TDAs. For close-in tactical operations the surface sensor data will be used. Determination of which weather data are used in a TDA is determined by the parameters of the operation, (i.e. proximity to surface sensors, date of mission, etc.) and will be invisible to the analyst. The analyst will, however, have the capability to override this function and select the weather data of his or her choice.

SUMMARY

The GOAL of the ALBE program is to provide battlefield commanders with improved war fighting capabilities. Generating products in minutes instead of the hours it now takes using manual methods is an important goal. The fusion of weather and terrain data into these products means the commander will have a product which automatically reflects the present or forecast situation on the battlefield - not the situation which occurred the day before. ALBE is working to provide Army Systems with the necessary tools to help soldiers fight and win on the battlefield of the future.

REFERENCES

Robert C. Brown, U.S. Army Atmospheric Sciences Laboratory, TEST WEATHER SYSTEM OVERVIEW, 1988

U.S. Army Atmospheric Sciences Laboratory, INTEGRATED METEOROLOGICAL SYSTEM (IMETS) PROOF OF CONCEPT PLAN (Draft), 1989

U.S. Army Atmospheric Sciences Laboratory, TEST AND EVALUATION PLAN FOR THE INTEGRATED METEOROLOGICAL SYSTEM (IMETS), 1989

DISCUSSION

T. FITZSIMONS, UK

Do you plan to deploy the ALBE system in NATO Europe? If so, how extensively and how will you satisfy requirements for data standardization to permit interoperability and data exchange?

AUTHOR'S REPLY

The AirLand Battlefield Environment (ALBE) Test-bed will never be fielded by the U.S. Army. The goal of ALBE is to provide software capabilities to Army systems that are in the field or are being developed for the field. The Digital Topographic Support System (DTSS) is one such system. DTSS will be fielded beginning in 1993 in Europe at the III, V, and VII Corps'. Its primary purpose is to update, generate and use digital terrain data; and to provide the capability of generating tactical terrain/weather graphics to the field commander. The format of the data which will be used by DTSS is that established by the Defense Mapping Agency's (DMA) Interim Terrain Data (ITD). DMA is represented on various international committees and organizations involved in the standardization and interoperability of terrain data between NATO members. I know of no efforts at standardization of weather data.

BATTLEFIELD OBSCURATION FACTORS

by
Melvin G. Heaps and Franklin E. Niles
U.S. Army Laboratory Command
Atmospheric Sciences Laboratory
White Sands Missile Range, New Mexico 88002-5501
USA

SUMMARY

Battlefield obscuration factors include the presence of man-made conditions as well as the natural atmosphere. An overview is given of the defeat mechanisms for electro-optical (EO) sensors likely to be encountered in the realistic battlefield environment. An example of friendly and threat obscurant usage is given to estimate the locations, amounts, and durations of smoke and obscurants likely to be found on the battlefield. Climate and adverse weather effects are factored in, particularly as they affect EO systems. In addition, the paper emphasizes the synergism between weather and battlefield obscurants. The general purpose is to provide the defense community with the data to quantify environmental and obscuration factors and their effects on EO systems. These include (1) providing data and methodology for system proponents and designers to assess the effects of battle-induced obscurants on EO components and systems, (2) providing the analytic community with information to calculate system performance, (3) indicating to the test and evaluation community the effects and levels of obscurants that should be considered when a system is evaluated, and (4) indicating to the training and doctrine community effects of realistic battlefield environments on system deployment and potential outcomes in such environments.

1. SYSTEM DEFEAT MECHANISMS

Smokes, obscurants, and adverse weather affect and, in turn, defeat electro-optical (EO), millimeter wave (MMW), and radar systems principally through the mechanisms of attenuation, scatter, backscatter, emission, contrast reversal, and clutter. Because most passive systems, and many active ones, rely on contrast differentials (or signal-to-noise ratios) to detect targets against the background, these defeat mechanisms can be thought of as ways to reduce, eliminate, or alter the target's contrast. Of course, there are many other types of defeat mechanisms. False targets, decoys, and emitters can all add clutter to the target scene. Jammers and active measures for target signature suppression or alteration can all lead to sensor defeat, without necessarily damaging the sensor itself. Smokes, obscurants, and adverse weather can affect these other countermeasures, but do not cause these countermeasures.

The task of sensors is to facilitate target acquisition. In addition to the sensor defeat mechanisms previously mentioned, there are several other factors that affect target acquisition, not all of which are obscurant or weather related. Table 1 provides a brief summary.

TABLE 1. FACTORS AFFECTING TARGET ACQUISITION

Atmospheric transmission	• Molecular extinction (particularly H_2O and CO_2 for infrared (IR) and H_2O and O_2 for MMW) • Aerosol extinction (haze, fog, and dust) • Precipitation • Battlefield smoke and dust
Path conditions	• Path radiance (scattered or emitted) • Refraction • Turbulence
Scene characteristics	• Intrinsic contrast • Changes in illumination • Clutter • Camouflage, concealment or deception • Target size and range • Target, moving or stationary • Target/background spectral content (color)
Imaging system characteristics	• Wide/narrow field of view • Detector sensitivity • Signal/noise (detection) threshold • Signal processing and display
Human characteristics	• Contrast threshold • Variability among operators • Variability within an operator

The emphasis in the past has been on essentially horizontal, near ground level, lines of sight (LOS). With attenuation considered as the primary defeat mechanism, several rules of thumb have been developed for sensor performance in battlefield obscuration in these situations. These are summarized in table 2.

TABLE 2. RULES OF THUMB FOR SENSOR PERFORMANCE IN SMOKES, OBSCURANTS, AND WEATHER

- All conventional smokes* and dust at anticipated battlefield levels will screen visible and laser designator systems on at least some portions of the battlefield.

- Phosphorus smokes and battlefield dust in large concentrations will screen thermal viewers (3-5 and 8-12 μm IR wavelengths).

- Conventional smokes other than phosphorus are not effective in the IR region unless large areas are screened so that LOS through these smokes are greater than 1 km.

- Specially developed IR screeners will screen thermal viewers in areas where they are employed; likewise, specially developed obscurants can screen MMW and radar systems. Visual systems may not always be screened, unless conventional smokes are also employed.

- Top attack weapons offer an advantage over horizontal (ground level) LOS weapons in battlefield smoke and obscurant environments.

- Smoke and obscurants, acting in concert with natural reductions in visibility, can be much more effective than when used alone.

- Natural reductions in visibility can defect visual sensors (day sights, direct view optics, day TV, etc.) and laser rangefinders.

- Thermal viewers provide an advantage over visual sensors in haze, fog, and light rain.

- In very thick fog, as well as snow, visual sensors and thermal viewers are both degraded in performance about equally.

- Haze, fog, light rain, and snow provide much less attenuation for MMW and radar systems than for visual or IR systems; however, medium to heavy rain will adversely affect MMW systems.

*Conventional smokes: hexaChloroethane (HC), fog oil, phosphorus, and anthracene

However, the emphasis is now shifting toward near vertical or large-angle slant paths because of the advent of top-attack weapons and increased possibilities for satellite and space-based observation systems. The next two sections consider the effects of natural, and in turn battlefield, obscurants.

2. WEATHER AND THE NATURAL OBSCURANTS

Any battle must be conducted against the backdrop of the regional climate, plus the daily weather. It is important to have an accurate feeling for the effectiveness of the natural obscurants, as well as for their frequency of occurrence. Natural obscurants, including the atmospheric gases, are almost always considered to have uniform distributions within the combat area, at least in the horizontal plane. A homogeneous path is generally assumed while modeling the optical effects of natural obscurants for horizontal LOS. Additional weather data, including frequencies of occurrence, are available in references 1 and 2.

Table 3 gives the volume extinction coefficients for natural obscurants at several wavelengths of interest; horizontal paths assume the obscurant is uniform along the entire path. Table 4 gives the approximate vertical optical depths as a result of the presence of natural obscurants, assuming one is looking down through the atmosphere from either high altitude aircraft or space-based assets (obscurant uniformity is not assumed along the path).

TABLE 3a. OBSCURATION EFFECTIVENESS AND APPROXIMATE RANGES OF EXTINCTION COEFFICIENTS (km^{-1})* FOR NATURAL OBSCURANTS IN THE VISIBLE, INFRARED, MILLIMETER WAVE AND RADAR REGIONS

Natural Obscurant	Spectral Region: Visible (0.4 to 0.7 μm)	Mid-IR (3 to 5 μm)	Far-IR (8 to 12 μm)
Gases	Very low* ~0.02	Low/medium 0.25 to 0.73	Very low/medium 0.03 to 0.8
Haze	Low/medium 0.2 to 2	Very low/medium 0.02 to 1	Very low/low 0.02 to 0.4
Fog	High 2 to 20	Medium/high 1 to 20	Medium/high 0.4 to 20
Clouds**	Very high 7 to 100	Very high 3 to 100	Very high 3 to 100
Rain	Low/medium 0.3 to 1.6	Low/medium 0.3 to 1.6	Low/medium 0.3 to 1.6
Snow	Medium/high 2 to 12	Medium/high 2 to 12	Medium/high 2 to 12
Dust	Low/high 0.2 to 4	Low/high 0.2 to 4	Low/high 0.2 to 4

Description	Extinction Coefficient (km^{-1})	Visual Range (km)
Very low	< 0.1	> 30, very clear
Low	0.1 to 0.5	6 to 30, clear to hazy
Medium	0.5 to 2	2 to 6, hazy
High	> 2	< 2, foggy

Table 3b.

Natural Obscurant	Spectral Region: MMW (94 GHz) (~ 3 mm)	MMW (35 GHz) (~ 9 mm)	Radar (10 GHz) (3 cm)	Radar (3 GHz) (10 cm)
Gases	Very low/low 0.03 to 0.2	Very low 0.02 to 0.06	Very low ~ 0.01	Very low ~ 0.003
Haze	Very low ~ 0.001	Very low < 0.001	-- ~ 0	-- ~ 0
Fog	Very low/low 0.01 to 0.04	Very low/low 0.001 to 0.1	Very low < 0.04	Very low < 0.01
Clouds**	Low/high 0.1 to 10	Very low/medium 0.01 to 1	Low < 0.4	Very low < 0.1
Rain†	Low/medium 0.3 to 2	Very low/medium 0.05 to 1	Very low < 0.1	Very low < 0.01
Snow‡	Low/medium 0.03 to 1	Very low/medium 0.0004 to 1	Very low < 0.01	Very low < 0.01
Dust	Very very low ~ 0.0005 to 0.005	Very very low 0.00005 to 0.005	-- ~ 0	-- ~ 0

*Many radar applications use attenuation in decibels/kilometers; to convert km^{-1}, multiply values in table by 2.3 (that is, $\log_e 10$).
**Stratus or fair weather cumulus clouds.
†Backscatter in rain is just as significant as extinction.
‡Extinction in snow depends greatly on the snow's liquid water content; the drier the snow the less the attenuation.

TABLE 4a. OPTICAL DEPTHS (TOP DOWN) FOR NATURAL OBSCURANTS IN VISIBLE THROUGH RADAR WAVELENGTHS

Natural Obscurant	Spectral Region: Visible (0.4 to 0.7 μm)	Mid-IR (3 to 5 μm)	Far-IR (8 to 12 μm)
Gases	0.1 to 0.2	1 to 3	0.1 to 3
Haze	0.2 to 2	0.02 to 1	0.02 to 0.4
Fog	0.2 to 5	0.1 to 5	0.05 to 5
Clouds*	2 to 20	0.3 to 20	0.3 to 20
Rain**,†	0.1 to 1.5	0.1	1.5
Snow**,‡	1 to 10	1 to 10	1 to 10
Dust	0.4 to 2	0.4 to 2	0.4 to 2

Table 4b.

Natural Obscurant	Spectral Region: MMW (94 GHz) (~ 3 mm)	MMW (35 GHz) (~ 9 mm)	Radar (10 GHz) (3 cm)	Radar (3 GHz) (10 cm)
Gases	0.2 to 0.8	0.1 to 0.25	< 0.1	< 0.02
Haze	~ 0	~ 0	~ 0	~ 0
Fog	< 0.2	< 0.02	< 0.01	~ 0
Clouds*	< 2	< 0.2	< 0.1	< 0.02
Rain**,†	< 2	< 1	< 0.1	< 0.01
Snow**,‡	< 1	< 1	< 0.01	< 0.01
Dust	~ 0	~ 0	~ 0	~ 0

*Stratus or fair weather cumulus (~ 200 m thick).
**Optical depth of cloud is in addition to value shown.
†For MMW and radar wavelengths, backscatter in rain is just as significant as extinction.
‡For MMW and radar wavelengths, extinction in snow depends greatly on the snow's liquid water content; the drier the snow the less the attenuation.

3. THE BATTLEFIELD: COMBAT OBSCURANTS

Weapon systems must be able to perform the task of target acquisition in an active, dirty, dynamic battlefield environment. In order to assess the performance of current and future systems in such environments, it is necessary to have an accurate knowledge of the anticipated battlefield obscuration levels. The purpose is to provide the defense community with data and a methodology to quantify obscuration factors and their effects on EO systems. Underlying these processes is a knowledge of the basic optical properties of smokes and obscurants, munition characteristics, dynamics of rapidly developing smoke clouds, transport and diffusion, physics, and meteorology.

Modules of the Electro-Optical Systems Atmospheric Effects Library (EOSAEL)[3] have been used extensively in preparing handbooks such as the Combat Environments Obscuration Handbook,[1] the Smoke and Natural Aerosol Parameters (SNAP) Manual,[4] and the Handbook for Operational Testing of Electro-Optical Systems in Battlefield Obscurants.[5] These handbooks quantify realistic battlefield environments, give examples of both threat and friendly smoke and obscurant usage, provide methods to simulate obscurants, and identify the instrumentation necessary to characterize the effects of obscurants. More detailed information can also be found in references 6 and 7.

The critical questions when smoke and obscurant effects are to be deployed are "Where will the smokes/obscurants be deployed?" "How long will they persist?" and "How much will there be?" The "where" question can be answered by considering where the EO system will be deployed and where the threat smokes/obscurants are projected. Not all obsurants will be present everywhere on the battlefield. The "how long" question gives rise to the rule of thumb that single events may last 10-30 min and large-scale, preplanned obscuration events may last from 30 min to over an hour. Large area screens in rear echelons may be designed to last several hours. Table 5 gives the mass extinction coefficients for the common battlefield obscurants. The anticipated ranges of amounts for both horizontal and vertical LOS are given in table 6.

TABLE 5. MASS EXTINCTION COEFFICIENTS (SQUARE METERS PER GRAM) FOR COMBAT-INDUCED OBSCURANTS

Obscurant	Spectral Region				
	Visible	Near-IR (1.06 μm)	Mid-IR (3 to 5 μm)	Far-IR (8 to 12 μm)	MMW/Radar*
Phosphorus**	4.08	1.37	0.29	0.38	0.001
HC	3.66	2.28	0.19	0.03	0.001
Oil-based	6.85	3.48	0.25	0.02	0.001
Anthracene	6.20	2.50	0.23	0.05	0.001
Vehicular and high explosive (HE) dust	0.32	0.26	0.27	0.25	0.01 to 0.001
Carbon	1.50	1.42	0.75	0.32	0.001

*Nominal values, these obscurants are essentially transparent at MMW and radar wavelengths.
**For 50-percent relative humidity at 10 °C.

4. SUMMARY

This paper has explored the concept of battlefield atmospherics, which deals with the presence of man-made as well as natural atmospheric conditions over potential battlefields, and the effects of realistic battlefield environments on materiel, personnel, tactics, or operations. The paper specially emphasizes the impact of the environment on EO sensors. The pertinent question for the military planner is whether the anticipated use of obscurants, coupled with the anticipated weather, as outlined previously, will exceed the defeat thresholds of the EO sensors. The following rules of thumb apply when estimating the effectiveness of smokes, obscurants, and the weather:

- Sometimes the obscured battlefield will exceed reasonable design and performance requirements.

  Tactics, not system design, then become paramount.

- Reasonable performance criteria for EO systems should require performance to an optical depth of 3 (that is, down to a one-way transmittance of 5 percent or a signal loss of up to 13 dB).

- Requirements for greater system sensitivity may become prohibitively expensive (that is, performance at optical depths of 5 or more, or one-way transmittance less than 1 percent, or a signal loss of more than 20 dB).

- Tests should be conducted to find system performance thresholds.

- Tests should be conducted to separate out defeat mechanisms (that is, find out precisely why the sensor or system failed).

Future directions should incorporate battlefield atmospheres into the materiel development process for both proponents and designers of future weapon systems. Realistic performance and design standards should be set for the EO components and systems that must function in battlefield environments. In addition, target acquisition models should incorporate battlefield atmospheres so that more realistic methodologies can be used to generate accurate probabilities of detection and acquisition, which are used to assess system performance. Finally, wargames should increase their consideration of battlefield atmospherics.

TABLE 6. OBSCURATION EFFECTIVENESS OF BATTLEFIELD SMOKES/OBSCURANTS[a] APPROXIMATE RANGES OF CONCENTRATION LENGTHS AND OPTICAL DEPTHS

		Horizontal Optical Depth[c,d] for Different Spectral Regions		
Battle-Obscurant	Concentration Length[b]	Visible IR (0.4 to 0.7 μm)	Mid IR (3 to 5 μm)	Far IR (8 to 12 μm)
Anthracene[e]	1 to 3	6 to 20	0.2 to 0.7	0.05 to 0.2[n]
Phosphorus[f] (Threat)	10 to 50	40 to 200	3 to 15	4 to 20
Phosphorus[g] (US)	2 to 10	8 to 40	0.6 to 3	0.8 to 40
HE dust[h] (Threat)	20 to 200	6 to 60	6 to 60	5 to 50
HE dust[i] (US)	1 to 10	0.3 to 3	0.3 to 3	0.2 to 3
Vehicular dust[j]	1 to 10	0.3 to 3	0.3 to 3	0.2 to 3
Diesel oil[k]	2 to 6	12 to 35	0.5 to 3	0.04 to 0.1[n]
Smoke grenades[l]	1 to 4	4 to 16	0.3 to 1	0.4 to 2[n]
Fog oil[m]	3 to 10	20 to 65	0.7 to 3	0.06 to 0.2[n]
		Vertical Optical Depth[c,d] for Different Spectral Regions		
Anthracene[e]	< 0.2	< 1	< 0.05	~ 0[n]
Phosphorus[f] (Threat)	3 to 10	12 to 40	1 to 3	1 to 4
Phosphorus[g] (US)	< 1	< 4	< 0.3	< 0.4
HE dust[h] (Threat)				
(small areas)	6 to 10	2 to 3	1 to 3	1 to 3
large areas)	1 to 3	0.3 to 1	0.3 to 1	0.3 to 1
HE dust[i] (US)	1	0.3	0.3	0.3
Vehicular dust[j]	< 0.5	< 0.2	< 0.2	< 0.2
Diesel oil/VEESS[k]				
(patches)	0.5	3	0.1	~ 0[n]
(clouds)	0.2	1	0.05	0
Smoke grenades[l]				
(v small patches)	1	4	0.3	0.4[n]
(clouds)	0.2 to 0.4	1 to 2	0.1	0.1
Fog oil[m]				
(small areas)	3	20	1	< 0.1[n]
(large areas)	1	6	0.3	~ 0

[a]Meteorological conditions are favorable for placing and maintaining the obscurant between the threat and friendly forward line of own troops (FLOT).

[b]The concentration length is the path integrated amount of obscurant (units of grams per square meter) along LOS from the friendly FLOT to the threat FLOT, or from an airborne platform to the surface.

[c]Anticipated optical depths along horizontal LOS between friendly and threat FLOTs, nominal 2 km separation, or from 100 m elevation to the surface.

[d]The optical depth (dimensionless) is the product of the concentration length times the obscurant's mass extinction coefficient appropriate for the given wavelength band.

[e]Many smoke pots contain this type of fill.

[f]Bulk-filled white phosphorus (WP) rounds; smoke clouds tend to pillar.

[g]US M-825 round; WP impregnated wedges; smoke clouds do not pillar.

[h]Conventional rounds filled with high explosive.

[i]Dual-purpose improved conventional munitions (produces less dust).

[j]Highly dependent on vehicle speed and state of soil.

[k]Used in vehicle engine exhaust smoke system (VEESS).

[l]Assumed red phosphorus fill.

[m]Used in smoke generators that are positioned behind the FLOT.

[n]Different fill materials can be used to increase IR screening properties.

REFERENCES

1. Combat Environments Obscuration Handbook (CEOH), Volume I - Natural and Combat-Induced Obscurants and Environments, 61 JTCG/ME-84-1-1, Joint Technical Coordinating Group for Munitions Effectiveness, 1986.

2. EOSAEL 87, Volume 8, Climatology Module: CLIMAT, ASL-TR-0221-8, U.S. Army Atmospheric Sciences Laboratory, White Sands Missile Range, NM, 1987.

3. EOSAEL 87, Volume 11, Combined Obscuration Model for Battlefield-Induced Contaminants: COMBIC, ASL-TR-0221-11, U.S. Army Atmospheric Sciences Laboratory, White Sands Missile Range, NM, 1987.

4. Smoke and Natural Aerosol Parameters (SNAP) Manual, 61 JTCG/ME-85-2, Joint Technical Coordinating Group for Munitions Effectiveness, 1985.

5. Handbook for Operational Testing of Electro-Optical Systems in Battlefield Obscurants, 61 JTCG/ME-87-10, Joint Technical Coordinating Group for Munitions Effectiveness, 1987.

6. Quantitative Description of Obscuration Factors for Electro-Optical and Millimeter Wave Systems, DOD-HDBK-178(ER), 1986.

7. Milham, Merril E., and David H. Anderson, Obscuration Sciences Smoke Data Compendium: Standard Smokes, ARCSL-SP-82024, U.S. Army Chemical Research, Development and Engineering Center, Aberdeen Proving Ground, MD, 1983.

A SCREENING SMOKE UTILIZATION TACTICAL DECISION AID

Mary Ann Seagraves and Melvin G. Heaps
U.S. Army Laboratory Command
Atmospheric Sciences Laboratory
White Sands Missile Range, New Mexico 88002
USA

SUMMARY

A screening smoke tactical decision aid (TDA) is being developed to help the battlefield commander choose which of his available smoke assets would be the most favorable for achieving his objectives for the expected tactical and environmental situation. Knowledge-based artificial intelligence techniques will be used to determine the relative ranking of the smoke assets based on a variety of tactical and environmental factors. TDA input will consist of a list of available assets and meteorological data for the time and area of interest. Information will also be input concerning the scenario, such as the intended use of the smoke and terrain-related information. A number of independent factors will be considered and each available smoke asset will be ranked as ideal, favorable, marginal, or unfavorable for each of these factors. An overall ranking will then be derived for each smoke type. TDA output will consist of the relative rankings of the assets, information concerning why the use of certain assets may not be favorable, and warnings concerning possible toxic effects on one's own troops.

1. INTRODUCTION

Obscurants are known to enhance force effectiveness and reduce vulnerability to enemy combat operations. U.S. forces will use obscurants whenever tactical advantage outweighs potential degradation to friendly operations. Use of smoke in the offense requires careful planning and execution to preclude interference with movement or assault operations, to retain the element of surprise, and to avoid silhouetting or drawing undue attention to friendly troops. Smoke use in the defense requires careful preparation to preclude an ill-conceived deception, disruption of friendly activities, or poorly timed low-visibility retrograde operations. The commander must be able to decide on types of obscurants to be employed depending on the capabilities of the delivery system, effects of weather and terrain, and the effects upon friendly and enemy target acquisition systems.

A screening smoke tactical decision aid (TDA) is being developed to help the commander choose which available smoke assets would be the most favorable for achieving his objectives for the expected tactical and environmental situation. The purpose of this TDA is to automate the process of determining how the environment might favor or adversely affect the use of particular smoke types and dissemination methods.

This TDA is being designed for use by the G3/S3 staff and chemical officers in assessing the environmental effects on smoke deployment and in planning for the optimum use of smoke assets. The TDA is expected to be fielded on the Integrated Meteorological System (IMETS) and will be demonstrated in the AirLand Battlefield Environment (ALBE) software demonstration program.

The approach being taken in the TDA development is to automate information available in field manuals, in handbooks, and from experts in the field. The automation process will utilize artificial intelligence (AI) techniques known as expert systems, which are programs that mimic the advice-giving capabilities of human experts.

2. EXPERT SYSTEM DESIGN

According to Brule (1) some fundamental questions that must be answered during the process of trying to make computer programs that exhibit AI are:

- How will knowledge be represented?
- What strategies will be used in arriving at answers to the problem?
- What will the sources of information to the system be?

According classical logic, questions such as "Is it raining outside?" must be answered "yes" or "no." There is no provision for an answer such as "sort of" or "not really." Then the yes (or true) is assigned a value of 1 and the no (or false) is assigned a value of 0. A set of conventions is used for translating logical propositions into mathematical equations. In the early 1960's an engineer named Lotfi Zadeh proposed a new approach, called fuzzy systems or fuzzy logic, as a method of representing intermediate values between true and false. Zadeh's method was to allow for an infinite range of values between 0 and 1 and to use a slightly different set of conventions for

translating logical propositions into mathematical equations. It is this approach that has been taken as appropriate for representing information and knowledge concerning the use of smoke assets.

Negoita (2) states that the power of expert systems lies in their knowledge bases and that moving knowledge from the brains of experts into computer programs is a process of working together with the experts. In the field of utilization of smoke assets, much knowledge has been documented in military field manuals and handbooks. However, there are many questions unanswered in the literature. To fulfill this need, a special task group is being formed within the Smoke and Aerosol Working Group (SAWG) of the Joint Technical Coordinating Group for Munitions Effectiveness (JTCG/ME). This task group is being charged with generating the information required by the TDA.

A number of factors or parameters have been identified as being important in the utilization of smoke. These factors are:

- function for which the smoke is to be used,
- type of sensor against which the smoke will screen,
- amount of time available before smoke is required,
- range to desired smoke location,
- atmospheric stability,
- air temperature,
- relative humidity,
- wind direction,
- wind speed,
- terrain type, and
- surface characteristics.

In the spirit of fuzzy logic, each parameter is assigned a ranking for each smoke type and dissemination method. These rankings and the corresponding numerical values are as shown in table 1. The rankings are stored in a data file which is accessed by the TDA. An example of parameter rankings for a few smoke types for atmospheric stability is shown in table 2.

TABLE 1. PARAMETER RANKINGS AND NUMERICAL VALUES EXAMPLE

Ideal	1.0
Favorable	0.7
Marginal	0.4
Unfavorable	0.0

TABLE 2. PARAMETER RANKING EXAMPLE FOR ATMOSPHERIC STABILITY

Smoke Type	Dissemination Method	Stable	Neutral	Unstable
Fog Oil	Generators	0.7	1.0	0.4
Diesel Oil	VEESS	0.7	1.0	0.4
HC	Artillery	1.0	0.7	0.4
WP	Artillery	1.0	0.7	0.4

The rules of inference, or the inference engine as it is known in AI circles, are critical. The task of the inference engine is to operate on the knowledge base in such a way that the knowledge is processed consistently and logically.

Some of the parameters relating to smoke utilization are likely to be more important than others. For example, there are situations in which the smoke toxicity is not a concern or the terrain type may not be known. A technique within the framework of fuzzy logic is used to weight the parameter rankings according to their presumed importance. First, an importance ranking of critical, important, marginal, minor importance, or negligible is assigned to each parameter. Then the parameter rankings are modified based on the parameter importance. An example of such a modification consists of linearly mapping the values of the parameter rankings from the interval [0,1] to progressively shorter intervals as the importance decreases from critical to negligible. These intervals, shown in table 3, each have 1 as their upper bound.

TABLE 3. IMPORTANCE RANKINGS AND MAPPING INTERVALS EXAMPLE

Ranking	Interval Length	Interval
Critical	1.0	[0,1]
Important	0.75	[0.25,1]
Marginal	0.50	[0.50,1]
Minor Importance	0.25	[0.75,1]
Negligible	0.0	[1,1]

Then, the modified rankings are combined for each smoke asset to give an overall ranking. Two methods are currently being investigated to derive an overall ranking. The first is to take the minimum value of the modified parameter rankings for each smoke type. The second is to take the nth root of the product of the modified rankings for each smoke type, where n is the total number of parameters, for example, 9. In either case, the result is a number in the interval [0,1] of the same order of magnitude as the modified rankings. Thus, if a smoke asset has a ranking of favorable (0.7) for all parameters and if all parameters are of critical importance, the overall ranking will be 0.7 for that smoke asset. A ranking of unfavorable (0.0) for a parameter of critical importance will result in an overall ranking of 0.0, regardless of how the other parameters are ranked for that smoke asset. The minimum value method for determining the overall ranking will result in a low ranking for a smoke asset if any of the modified parameter rankings for that asset are low. The product method will result in a higher overall ranking even if some parameters are ranked low (but not 0) but most of the parameters are ranked high.

The ranking for each smoke asset is then inferred from the overall numerical rankings converted to words. An example of this conversion is shown in table 4. These word rankings for each available smoke asset are the primary output of the TDA.

TABLE 4. OVERALL NUMERICAL RANKINGS EXAMPLE

Ideal	0.8 - 1.0
Favorable	0.55 - 0.8
Marginal	0.3 - 0.55
Unfavorable	0.0 - 0.3

3. TDA FEATURES

The TDA is designed to use meteorological data from a meteorological data file, such as available in IMETS and the ALBE testbed computer. A stand-alone version will be available in which the user will be required to input meteorological information. Other data required of the user are:

- available smoke types and dissemination methods,
- function for which the smoke is to be used (for example, camouflage or decoy),
- type of sensor against which the smoke is to screen (for example, thermal imager or laser),
- amount of time available until smoke is required (immediate, quick, or preplanned)
- range to location where smoke is desired (0-1 km, 1-5 km, or > 5 km),
- terrain type (flat, average, or rugged), and
- surface characteristics (wooded, open, or urbanized).

The user will be able to delete smoke types and dissemination methods from the menu list. He will be able to change the parameter importance rankings. A version will be available in which the user will be able to add smoke assets to the menu and change the parameter rankings, but this feature will probably not be available in the fielded version of the TDA.

The user will also be able to determine which parameters contribute to the overall ranking of a particular smoke asset to cause it to be rated as marginal or unfavorable.

4. CONCLUSIONS

A TDA now being developed will give the battlefield commander information concerning which of his smoke assets will be the most favorable to use in a particular situation by considering environmental effects. An initial prototype program has been written to use in developing the final knowledge base. FORTRAN was used for this

initial prototype and will probably be used in the final TDA. This language was chosen instead of one of the AI-oriented languages because it will be available on the targeted systems, while the AI languages probably will not be.

The final knowledge base is being developed. After its completion the final TDA program will be written. A finished product is expected to be available by September 1989.

REFERENCES

1. Brule, J. F., Artificial Intelligence Theory, Logic and Application, Tab Books, Inc., Blue Ridge Summit, PA, 1986.

2. Negoita, C. V., Expert Systems And Fuzzy Systems, Benjamin/Cummins Publishing Co., Menlo Park, CA, 1985.

AIR FORCE TACTICAL DECISION AID DEVELOPMENT

James Hancock, Major USAF/Dr. Michael J. Kraus

Geophysics Laboratory
Hanscom AFB, MA 01731
USA

SUMMARY

Precision Guided Munitions (PGMs) and Target Acquisition Systems (TASs) allow greater stand-off ranges, increased accuracy, and higher kill probabilities than conventional bombing methods. They are also more sensitive to the environment. This increased environmental sensitivity led to the tasking of Air Weather Service, by the Tactical Air Forces, to provide weather support for PGMs and TASs. This support is based on Tactical Decision Aids (TDAs) that have been developed by Air Force Systems Command laboratories for Air Weather Service.

Conventional weather support is not sufficient when forecasting for "smart" weapons. Target detection and recognition depends on the target and background characteristics, the intervening atmosphere, and the sensor characteristics. The TDA models these three factors for infrared (8-12 micrometer), TV, and 1.06 micro-meter laser designator systems.

The TDA has been provided to Air Weather Service for operational use, but AFSC continues to produce upgraded versions with improved physics and expanded target, background, and sensor menus. Experience with the TDAs has been good, they have proven to be an effective, useful tool for planning and executing Air Force missions; but the challenge remains to provide the best, most effective tool possible, and to provide the most accurate and up-to-date environmental information in the target areas. The Air Force is addressing this requirement too.

1. INTRODUCTION

Soon after precision guided munitions and target acquisition systems were introduced into the Air Force weapons inventory, Tactical Air Forces identified a need for expanded weather support. Aircrews needed forecasts of target-to-background scene contrast and effective range of their weapon system to successfully plan and execute their missions. Elements determining mission success included information about the target and background physical characteristics, the influence of weather on the target scene, atmospheric transmittance, sensor performance characteristics, and attack geometry. Air Force Systems Command (AFSC) has developed an electrooptical tactical decision aid (TDA) program that considers these elements to provide scene contrast and range forecasts needed by aircrews. These operational TDAs (OTDAs) run on microcomputers used by Air Weather Service (AWS) forecasters supporting missions involving infrared (8-12μm), television, and 1.06μm laser weapon systems.

The Air Force TDA program is well structured, recognizing the changing needs of AWS who, in turn, provides the TDA forecast support to the Tactical Air Forces (TAF), the Strategic Air Command (SAC), and the Military Airlift Command (MAC). These requirements from AWS drive the development and model improvement efforts conducted by AFSC.

AFSC has not worked in isolation during the electrooptical TDA development. There is a continual exchange of technology between AFSC, the Army Atmospheric Sciences Laboratory, and the Naval Environmental Prediction Research Facility (NEPRF). Additionally, groundwork has been established for technical exchange between AFGL and the Army Cold Region Research and Engineering Laboratory's (CRREL), target scene simulation project called Balanced Technology Initiative (BTI).

2. TDA TEAM MEMBERS AND THEIR ROLES

Air Force Systems Command (AFSC) tasked the Geophysics Laboratory (GL) to be the central figure in the development of the TDA program. Additionally, AFSC tasked the Wright Research and Development Center (WRDC) to assist GL by managing the development and assessment of a research grade TDA (RGTDA). Research grade modeling is achieved without the size and run time constraints imposed by the mission requirements on the OTDA. Output from the RGTDA is assessed against measured data to identify and correct deficiencies in the code, and to improve modeling techniques. Completed RGTDA programs are delivered to GL for production of the OTDA that is delivered to AWS and distributed to forecasters supporting training and operational missions. Feedback from AWS users and the aircrews they support funnels back through AWS headquarters to GL establishing requirements for model development and improvement.

3. SOFTWARE STRUCTURE

Each of the TDAs has a modular construction with three sections: a target model, an atmospheric transmittance model, and a sensor model.

3.1 INFRARED

The IR target model computes target and background temperatures to produce their inherent thermal contrast. Radiant emittance for the spectral region of the sensor is calculated from the computed temperatures. The transmittance model uses Lowtran 5, 6, or 7 to compute transmittance over a four kilometer range and Beer's Law is used to approximate transmittances over other ranges. The Mark III version contains two new aerosol transmittance models, desert and Navy maritime. The sensor performance model calculates ranges based upon the particular specifications for the sensor employed.

3.2 TELEVISION

The TV model calculates contrast between the target and background using reflected visible or near-infrared radiation. For each TV sensor in the OTDA, transmittance is computed from Lowtran using a

shortened wavelength interval that is representative of the broader wavelength interval of the sensor. This process shortens OTDA r· ·ime with little effect on model accuracy. As with the IR model, the range calculation is based on unique characteristics of the sensor.

3.3 LASER

The laser target model computes the amount of laser energy reflected into the view of the sensor. Atmospheric extinction is determined using Lowtran, even though Lowtran is a broadband model as compared to the laser wavelength spread. This simplification to transmittance calculation is possible because the only significant extinction mechanism over the Lowtran wavelength interval bounding 1.06 μm is aerosol extinction which is nearly constant over the interval. As with other approximations in the OTDA, this shortens run time with only a small affect on accuracy. The laser range model produces target designator or receiver range for most laser systems. A new capability with the Mark III version forecasts the maximum effective distance that laser ranging systems may be employed.

4. MODEL IMPROVEMENTS

4.1 HIGH VALUE TARGETS

TDAs were originally developed to support missions against targets of opportunity, primarily tanks. In fact, all of the targets modeled in the current IR OTDA are vehicular targets. As the utility of the TDAs became realized, requirements evoled for expanded targeting to support interdiction missions. The earliest requirements for these "high-value targets" are listed in Table 1; the first five targets on the list are incorporated in the Mark III OTDA and the rest are under development in the RG TDA. Future modeling efforts will expand the list further.

TABLE 1. HIGH VALUE TARGETS

Dam
POL Tank
Suspension Bridge
Bunker
Power Plant
Factory
Aircraft
Ship

4.2 FACETED TARGETS

Faceted target models, new in the Mark III OTDA, simulate the geometry of the target. This provides a more accurate representation of the cross-sectional areas of the target as viewed by the sensor. Also, heating and cooling of each facet is treated individually, which permits a better representation of heat transfer between the target and its environment, and better simulation of radiative properties of the target. This also permits modeling of 'hot spots' (or cold spots) on the target. If the area of the facet is large enough, and the facet-to-background temperature contrast is high enough, the model will produce an acquisition or lock-on range for the facet. This should produce a longer, more realistic range forecast, particularly with facets that produce heat.

4.3 GENERIC TARGETS

The general physical properties of any particular target within a class of that target will not vary by a great extent. For example any T-62 tank should have dimensional, thermal, and radiative characteristics similar to any other T-62 tank. But with the target list expanding to high value targets the theory of commonality of physical properties among the targets within a class can no longer be generally applied. For example, a physical model of one lock and dam system may not be usable to predict IR weapon effectiveness against another lock and dam system. We have cautioned our users to this fact. Nevertheless, many engineering aspects of classes of high value targets can be generalized and the physical characteristics of those aspects can be treated "generically". As an example, a class of bridges may be constructed of concrete highway supported by painted steel I-beams and concrete pillars; the thermal and radiative properties of concrete and painted steel can be included in the model and treated generically. Thus high value targets can be modeled in the IR TDAs as long as a mechanism exists for the user to input some of the physical characteristics of the particular target for a given air strike. In the bridge example, user input may include target dimensions, concrete thickness, and color of paint. Our "test case" generic target offered in the Mark III OTDA is the POL tank, where the user has the option to input the tank's height, diameter, fullness, and color. A more complex model is the generic building being developed in the RGTDA. An important part of the development process of high value targets is model assessment. Feedback from assessments is used to identify weaknesses in the model, which are, then, examined and remedied.

4.4 BACKGROUND MODELS

The Mark III version will contain three general, first principles background models: sand, water, and vegetation. Other general background types being developed in the RGTDA include snow, concrete, and asphalt. Work will soon begin to develop a research grade sand-soil "hybrid" background with user input variables of percent mixture and moisture content.

4.5 CLUTTER

Even though thermal contrast may be good, clutter in the scene will reduce acquisition range. To account for this, algorithms based on the scene complexity and thermal contrast have been developed and are included in the IR OTDA.

4.6 BATTLE INDUCED CONTAMINANTS (BIC)

BIC can be introduced into the target scene in three ways with each effect treated individually. First, small dust particles can become suspended over the target area and will remain suspended for a long period of time. These aerosols reduce atmospheric transmittance, thus affecting contrast and shortening range. This type of BIC is modeled in the IR OTDA; the TV and laser OTDAs will be improved to account for this battlefield haze. Secondly, large particles of dust from explosions can obscure parts of the battle arena for short periods of time. These plumes can temporarily nullity the effectiveness of electrooptical weapon systems when multiple passes are made against a target. To assist mission planners in the timing sequence of attacks, rules-of-thumb are being developed to determine fallout rates of debris raised by bombs. Finally, smoke and dust obscurants can be used as camouflage; the smoke and dust types listed in the OTDAs are being updated to reflect current technology. Army research in these areas will prove very useful.

5. SUMMARY

The Air Force electrooptical TDA program is continually being improved to meet the needs of AWS and their customers. The product is being used to support training and operational missions worldwide. Post mission analysis is proving over and over that aircrews who use TDA forecasts during mission planning are experiencing quicker target acquisition, less exposure time in the threat area, and a higher percentage of successful attacks.

AIRBORNE FLIR DETECTION OF SURFACE TARGETS

Herbert G. Hughes
Ocean and Atmospheric Sciences Division
Naval Ocean Systems Center
San Diego, California 92152-5000

SUMMARY

An algorithm is presented for predicting the detection ranges of a surface target by an airborne Forward Looking Infrared (FLIR) system. The total infrared background radiance scene under cloud-free skies is modeled to include the atmospheric path emissions between the target and sensor and the effects of a wind ruffled sea on the surface emissions and sky radiance reflections. A model is also introduced of the average temperature of a ship based upon the solar heating effects throughout a specified course, the ambient meteorological conditions, and the viewing angle. Together, these two models allow the range to be determined where the difference between the apparent ship's temperature (i.e., the actual ship temperature degraded by the atmospheric transmittance) and the effective background temperature of the sea surface as viewed from the sensor altitude is equal to the minimum detectable temperature difference of the FLIR. A case study is presented to demonstrate the vulnerability of a Frigate class ship to detection by an airborne common module FLIR during a five hour period where the ship's course changed allowing solar heating of different sides of the ship. The results of this study show considerable increases in predicted detection ranges with altitude using the present algorithm over those based on a fixed temperature difference between a target and its background.

INTRODUCTION

The stand-off ranges at which an adversary can detect and/or track a surface ship using passive infrared (IR) sensors is of primary importance for a ship commander to be able to estimate the times allowable for evasive actions against guided weaponry launched at the ship or for the deployment of countermeasures. Algorithms which are presently operational in the U. S. Navy [1,2] for predicting the performance of an airborne Forward Looking Infrared (FLIR) system operating against a surface target are based upon a fixed temperature difference between a target and its natural background. These algorithms determine the range at which the assumed temperature difference is degraded by the atmospheric infrared transmittance to the minimum detectable temperature difference of the sensor system. This approach neglects the effects of a wind ruffled sea on the surface emissions and sky radiance reflections and the atmospheric path emission contributions to the total background radiance scene which changes with viewing angle and altitude of the sensor. In this paper, a FLIR detection range algorithm is introduced which includes (1) the three aforementioned contributors to the sea radiance under cloud-free skies to derive (with the Cox-Munk [3] sea surface wave slope statistical model and the LOWTRAN 6 [4] computer code) a background model which varies with sensor altitude and viewing (zenith) angle, and (2) a model of the average ship temperature which is based upon solar heating effects throughout a specified course and the ambient meteorological conditions. Together, these two models allow the range to be determined where the difference between the ship's apparent temperature (i.e.,the actual ship temperature degraded by the atmospheric transmittance) and the effective background temperature as viewed from the sensor position is equal to the minimum detectable temperature difference of the FLIR. A case study is presented (using an actual course of a Frigate operating off the coast of San Diego, California) to demonstrate the use of the algorithm as a Tactical Decision Aid (TDA) to predict the vulnerability of a Frigate class ship to detection by an airborne common module FLIR. During a five hour period the ship's course changed allowing solar heating of different sides of the ship. Supportive airborne FLIR measurements were not available to assess the accuracy of the detection range predictions. However, airborne measurements of meteorological parameters and the sea surface temperature (obtained in the ship's operating area) and surface based measurements of 8 to 12 μm infrared sky/sea radiances were used to select and evaluate the background model used in the detection range predictions. As radiative temperature measurements onboard the ship were also not available for comparison with the modeled values, the predicted average temperature for the ship could only be compared to that measured (with corrections for atmospheric effects being taken into account) when the ship passed within about 1.7 km of a calibrated infrared imaging system near the end of the five hour cruise at sea. In the following sections the background radiance model, and the ship temperature calculations used to predict the detection range envelopes for the selected ship's course are presented for different altitudes of the FLIR. These detection ranges are then compared to those calculated using a constant temperature difference between the ship and the sea background.

MATHEMATICAL FORMULATION OF BACKGROUNDS

Consider the atmosphere to be composed of a number, n, of isothermal layers characterized by temperature T_i and transmittance $\tau(\nu,i,\mu)$ along the optical path traversing the ith layer at angle μ, and ν is the spectral wave number. From Kirchoff's law, the radiance of the ith layer is

$$N(\nu,i,\mu)_{sk} = [1 - \tau_a(\nu,i,\mu)]W(T_i)/\pi \qquad (1)$$

where $\tau_a(\nu,i,\mu)$ is the absorption transmittance and $W(T_i)$ is Planck's blackbody radiation formula. Then the spectral radiance reaching the sea surface through the intervening atmosphere is

$$N(\nu,i,\mu)_{sk}[\prod_{j=1}^{i-1} \tau(\nu,j,\mu)] = [1 - \tau_a(\nu,i,\mu)][\prod_{j=1}^{i-1} \tau(\nu,j,\mu)]\, W(T_i)/\pi \qquad (2)$$

Summing the contribution from all layers, the spectral radiance at the sea surface is then

$$N(\nu,\mu)_{sk} = \sum_{i=1}^{n} \{[1 - \tau_a(\nu,i,\mu)][\prod_{j=1}^{i-1} \tau(\nu,j,\mu)]W(T_i)/\pi\} \quad (3)$$

As shown in Figure 1, the radiance is allowed to strike a wave facet on the ocean surface with a gaussian distribution [3] of angular tilts α and β in the up-wind and cross-wind directions, respectively, such that an amount $N(\nu,\mu)'$ is reflected into the sensor at an altitude H_i within the mth layer. The probability that radiance hits the facet is equal to the probability that the wave slope exist, i.e.,

$$N(\nu,\mu)' = P(S_x,S_y)N(\nu,\mu)_{sk} \quad (4)$$

where

$$P(S_x, S_y) = 1/(2\sigma_x\sigma_y)EXP[0.05(S_x^2/\sigma_x^2 + S_y^2/\sigma_y^2)] \quad (5)$$

and $S_x = \tan\alpha$, $S_y = \tan\beta$, $\sigma_x^2 = 0.003 + 1.92x10^3V_c$, $\sigma_y^2 = 3.16x10^3V_c$, with V_c being equal to the current wind speed in the azimuthal direction ϕ with respect to the sensor. Then the total spectral radiance that is reflected from all the wave facets into the line-of-sight of the detector located in the mth layer is

$$N(\nu,\theta)_{rsk} = [\prod_{j=1}^{m-1} \tau(j,\theta)]\sum_{\mu} R(\nu,\Omega)P(S_x,S_y)N(\nu,\mu)_{rsk} \quad (6)$$

where $R(\nu,\Omega)$ is the complex reflectivity of sea water at the reflection angle Ω. In the above equations, both μ and Ω are implicit functions of S_x and S_y given by [5]

$$\cos\mu = (2S_x/A)\cos\theta'\cos\phi + (2S_y/A)\cos\theta'\cos\phi - (B/A)\sin\theta' \quad (7)$$

$$\cos\Omega = (S_x/A)\cos\theta'\cos\phi) + (S_y/A)\cos\theta'\cos\phi + (1/A)\sin\theta' \quad (8)$$

where $A = S_x^2 + S_y^2 + 1$, $B = S_x^2 + S_y^2 - 1$ and θ' is the sensor's zenith angle at the sea surface reflection point.

In a similar manner, the spectral radiances emitted by the sea surface wave facets (N_{ss}) and the path radiance (N_p) which reach the sensor at the zenith angle θ are are given by

$$N(\nu,\theta)_{ss} = [\prod_{j=1}^{m-1}\tau(\nu,j,\theta)]\sum_{\Omega} P(S_x,S_y)[1 - R(\nu,\Omega)]W(T_{ss})/\pi \quad (9)$$

and

$$N(\nu,\theta)_p = \sum_{i=1}^{m} [1 - \tau_a(\nu,i,\theta)][\prod_{j=i+1}^{m-1} \tau(\nu,j,\theta)]W(T_i)/\pi \quad (10)$$

where T_{ss} is the sea surface temperature, and again, the angle μ is implicit in the reflection angle Ω.

Then, the total spectral radiance reaching the detector is the sum of the three components

$$N(\nu,\theta)_{tot} = N(\nu,\theta)_{rsk} + N(\nu,\theta)_{ss} + N(\nu,\theta)_p \quad (12)$$

The total spectral radiance must then be averaged over the response of the FLIR system which in this case is taken to be the 8-to-12 μm wavelength band. Subroutines have been introduced in to LOWTRAN 6 [5] to calculate the total band averaged radiance. The reflection and zenith angles are calculated with equations 5, 7 and 8 corresponding to the incremented values of wave slopes in the intervals $-3\sigma_{x,y} < S_{x,y} < 3\sigma_{x,y}$. To limit the number of calculations, the zenith angles are divided into a maximum of 30 classes (with the criterion that each class should contain at least 10% of the probability), and the averaged angles for each class are then used in the radiance calculations.

CALCULATIONS OF BACKGROUND RADIANCE SCENE

For this study, a Piper Navajo aircraft, equipped with Rosemount temperature and pressure probes, and an EG&G dewpoint sensor, made a vertical spiral over the ocean to obtain the profile of temperature, relative humidity an pressure which are required inputs to the LOWTRAN 6 computer code for calculating the sea and sky radiances. A Barnes PRT-5 radiation thermometer was also onboard the aircraft for measurement of the sea surface temperature from low level constant altitude flights. The vertical profiles of temperature and relative humidity, which were measured at 1330 PST approximately 9 km off the coast of San Diego, California, are shown in Figure 2. The sea surface temperature measured during this period was 16.4 °C. The profiles extending up to an altitude of 2700 m were divided into 33 layers as allowed by LOWTRAN 6. The lower layers of the profiles are also divided into sublayers containing the same amount of absorbing and scattering material and the temperature as the original layer. This artificial layering has been found necessary [6] to remove the anomalous dip [7] which occurs when aerosols are included in the LOWTRAN 6 radiance calculations for zenith angles close to 90°. The LOWTRAN 6 aerosol model chosen for the calculations is the Navy Maritime Aerosol Model. This model is the sum of three lognormal size distributions and, in addition to the surface wind speeds (current and 24 h averaged) and relative humidity, requires the input of an air mass factor which identifies the origin of the aerosols as either marine or continental and is allowed to range between integer values of 1 for open ocean to 10 for coastal regions. Also, when an observed surface visibility is available as an input, the model is adjusted so that the calculated visibility at a wavelength of 0.55 μm is the same as the observed value. The air mass factor is defined in terms of atmospheric radon content or an air mass trajectory analysis to determine the time the air mass has been over land. As neither of these techniques were available, an alternate method was used to select an appropriate air mass factor. Near the time that the meteorological parameters were obtained, measurements of IR (8 to 12 μm) horizon

radiances were made with a calibrated thermal imaging system (*AGA THERMOVISION, model 780*) using a 2.95° field of view lens with an instantaneous field of view of 0.87 mr. The response of the system was determined by placing a blackbody of known temperature (±0.1°C for temperatures <50°C) in front of the lens aperture. The digitized video signal transfer function of the system then allowed the blackbody temperature to be reproduced to within ±0.2°C. For these measurements the scanner was located at an elevation of 30 m on the Point Loma peninsula in San Diego and was directed in a southerly direction over the ocean such that approximately half of the field of view was above and half below the horizon. The data processing software of the AGA system allows the effective blackbody temperature of each pixel in the scene to be displayed on the computer terminal screen. The pixel corresponding to the maximum temperature in the scene is taken to coincide with the infrared horizon. Using the current and 24 h averaged wind speeds ($V_c = 2.9$ m/s and $\bar{V} = 2.8$ m/s) measured on shore and the vertical profiles of meteorological parameters measured by the aircraft, LOWTRAN 6 calculations were made to agree with the maximum pixel radiance (16.5 °C or 3.23 mW/cm^2 sr) in the scene using nonunique combinations of air mass factors and visibilities. (It should be noted that these calculations were made using a modified current wind speed component, $A_3 = 10^{(0.06V_c - 2.8)}$, which is different from the value published in LOWTRAN 6. This modification was found to be necessary in order to match previously published measurements of IR sky radiances and near surface aerosol size distributions [7] using the model). As the AGA scanner could not be accurately plumbed, the zenith angle of the infrared horizon was taken be 0.01° less than the angle for which the LOWTRAN calculations indicated the refracted ray path first hit the earth. In this case the zenith angle corresponding to maximum radiance is 90.17°. In Figure 3 the solid line represents the locus of points which allows the LOWTRAN calculations to match the measured horizon pixel radiance with the different combinations of air mass factors and visibilities. At the time of the measurements Los Coronodos coastal islands off San Diego were barely visible to the naked eye at ranges between 25 and 35 km. In the figure, the integer values of 3 and 4 correspond to visibilities close to these ranges of 23 and 37 km, respectively. Figure 4 shows the comparison of the measured and calculated IR radiances for zenith angles within about 1° above and below the horizon using an air mass factor of 3 and a visibility of 37 km. Both the calculated sky (θ < 90.17°) and sea (θ > 90.17) radiances are in good agreement with the measured values for this low wind speed case. Whether or not similar agreements can be obtained for higher wind speed conditions needs to be determined.

Using the selected atmospheric model, the contributions of the path, sea and reflected sky radiances to the total background radiance were calculated as a function of altitude and zenith angle. In Figure 5, an example is presented of the calculations for a sensor altitude of 2000 m. For zenith angles less than about 95°, it is interesting to note that the major contribution to the background is the path emission with the reflected sky radiance being less than 10% of the total. These relative contributions will of course change for other altitudes. In Figure 6 the total apparent blackbody temperature of the sea background from the three contributors is plotted versus zenith angle for sensor altitudes of 500 m and 2000 m. At the 500 m elevation, the dip in temperature at about 97° is a result of the rapid fall off of path emission with increasing zenith angle (i.e., shorter slant paths to earth than at the 2000 m). For zenith angles greater than about 100°, there is little difference in the apparent temperatures at each altitude and both approach the measured sea surface temperature at the nadir zenith angle.

AVERAGED SHIP TEMPERATURE MODEL

The computer code SHIPSIG [8] for predicting the average temperature of a surface vessel was developed at the Naval Surface Weapons Center. The original BASIC version of the code has been rewritten in FORTRAN language for the HP-9020 computer. Basically, the model approximates the complex structure of a ship with a single plane element which represents the ship on an average basis. For a given viewing direction, the simplest representation of a ship consists of a single vertical element and a horizontal element with the observer's orientation accounted for by appropriate area components. The infrared signature calculations are then based on a thermodynamic analysis of both elements individually. They are combined by scaling the element radiance in proportion to the ship area each represents. The thermodynamic properties of the horizontal and vertical elements and ship stack correction factors applied to the vertical element presently listed in the SHIPSIG code are for a guided missile frigate class ship. The model requires as inputs the ship's course and speed as a function of time from a starting geographic latitude, the surface wind speed and direction, visibility, relative humidity, air temperature, the ship's initial temperature, and the viewing angle.

Figure 7 shows the course of a guided missile frigate, the USS BROOKE (FFG1), off the coast of San Diego, California on 9 June 1988, which was chosen to demonstrate the model. During the five hour time period, changes in the ship's heading allowed solar heating of different sides of the ship. As the ship completed the course and returned to harbor, it passed close to the the AGA thermal imaging system located on shore about 2 km from channel buoy #6 near the entrance to the harbor. The AGA system's data processing software allows subtraction of the sea background radiance surrounding the ship and provides a histogram of the temperature distribution of the ship pixels within a chosen rectangular area. In this case, the mean temperature (uncorrected for atmospheric effects) was 19.7°C, and the temperature distribution approximated a Gaussian curve remarkably well.

The measured radiance, N(meas), of the ship at a range R is related to its actual effective blackbody radiance, N(ship), and the atmospheric emission, N(path), along the path by

$$N(meas) = N(ship)\tau(R) + N(path) \qquad (13)$$

where $\tau(R)$ is the atmospheric transmittance at a range R. The range to the ship was determined to be approximately 1.7 km using the known vertical dimensions of the ship and their angular subtense within the field of view of the AGA. The relative humidity (72%), air temperature (20°C) and pressure (1012.4 mb) measured at the AGA location were used in LOWTRAN 6 calculations of transmittance and path emission to determine the temperature equivalent to N(ship). Figure 8 shows the adjusted temperature dependence on visibility and air mass factor. Conveniently, both of the combinations of air mass factor and visibility (AM = 3, Visibility = 37 km and AM = 4, Visibility = 23 km) result in the same adjusted ship's temperature of 20.5°C.

For the model calculations, the initial position of the ship was taken to be near the entrance to San Diego harbor. The initial ship temperature, its ambient temperature and relative humidity throughout the course were not recorded by the ship. The relative humidity was then taken to be constant at as measured at the AGA site. The surface wind was southwesterly at 2.9 m/s and the visibility was taken as 37 km. The depression angle of viewing was 0.6°. The average ship temperature calculated for the port side of the ship as a function of time is shown in Figure 9 assuming the initial ship temperature to be equal to the indicated ambient air temperatures which remained constant throughout the cruise. The most apparent feature is the ship's temperature response to the gradual heating of its port side as it steamed westward in the early morning, and the abrupt cooling after 1000 hours following the southeasterly course change at 0952 hours. While the shapes of the response curves do not appear sensitive to the uncertainties in ambient air temperature, their magnitudes differ by amounts equivalent to the uncertainties. Allowing for the uncertainties in the meteorological parameters surrounding the ship throughout the course, the reasonable agreement between the adjusted AGA temperature measurements near 1345 hours and the model is gratifying.

DETECTION RANGE CALCULATIONS

The concept of maximum detectable range (MDR) calculations using a fixed difference between a target at a temperature, T_s, and its effective background temperature, T_b, is depicted in Figure 10. The maximum detectable range is defined as that range where the actual temperature difference (T_s - T_b) is degraded by the atmospheric transmittance, $\tau(R)$, to an apparent temperature difference, ΔT_a, equal to the minimum detectable temperature difference (mdtd) of the system. As stated earlier, this approach neglects effects of a wind ruffled sea on the surface emissions and sky reflections and the atmospheric path emissions which contribute to the total background scene under cloud-free skies. It also neglects the path radiance between the target and detector which must be accounted for in a temperature difference concept for detection range predictions. The radiance contrast between a ship and its background can be converted an equivalent temperature difference if the slope of the radiance gradient with temperature is specified, i.e.,

$$(N_s - N_b)/(T_s - T_b) = \partial N/\partial T_{T=T'} \qquad (14)$$

or,

$$\Delta T_{eq} = \frac{\Delta N}{\partial N/\partial T_{T=T'}} \qquad (15)$$

where the wavelength dependency of N is implied. The equivalent temperature difference is somewhat sensitive to the choice of the temperature, T', at which the gradient is chosen. For the 8 to 12 μm wavelength band, the gradient at 280°K is approximately 11% lower than that at 290°K, and 27% at 310°K [9].

In the calculations presented here, the unmodified LOWTRAN 6 code is used to directly calculate the sum of the ship and path *radiances received by the sensor* at a range R as

$$N(R)_{s+p} = N(R=0)_s \tau(R) + N(R)_p \qquad (16)$$

$N(R)_{s+p}$ is then converted to an equivalent blackbody temperature $T(R)_{s+p}$ by an iterative solution to Planck's blackbody formula. Similarly, an equivalent blackbody temperature, $T(R)_b$, of the background radiance (Equation 12) at the specified range is calculated and the resulting apparent temperature difference , $\Delta T(R)_a = T(R)_{s+p} - T_b(R)$, determined. As shown in the inset of Figure (10), the intersection of the curve of $\Delta T(R)_a$ plotted vs range with that of the system's minimum detectable temperature difference (mdtd) curve determines the maximum detectable range (MDR) of the ship. The FLIR system mdtd versus range (spatial frequency) curve was calculated for a hypothetical FLIR operating against a rectangular target. In Figures (11) and (12), the calculated MDR's for the USS BROOKE by an airborne FLIR operating at altitudes of 0.5 km and 1.0 km, respectively, are shown. The MDR's could have been calculated as a function of time throughout the the ship's course, however, for the sake of simplicity, only the vulnerability detection envelopes for the entire duration of the ship's course are shown. In the figures, the ship's average temperature for both the port and starboard sides are shown. In Figure 11, the ship is seen to be vulnerable to infrared detection throughout its course from an altitude of 0.5 km at a range of 31 km. However, beyond 35 km the ship is not detectable. Similarly, in Figure 12, the ship is vulnerable at a range of 53 km from an altitude of 1.0 km but safe from but safe from detection at 56 km. In these examples, it is interesting to note that the temperature responses of the different sides of the ship to solar heating follow the course changes remarkably well. In Figure 13 a comparison is shown of the predicted detection ranges using the present algorithm and those assuming a constant temperature difference between the ship and its background of 5°C [1]. Considerable increases (≈20 km at an altitude of 2.0 km) in predicted detection ranges with altitude are obtained using the present method over the fixed-temperature method.

CONCLUSIONS AND RECOMMENDATIONS

The results of this case study have shown the reliability of the sea surface radiance model to accurately represent measured values for a low wind speed condition. Whether or not it will be representative of other wind speed conditions needs to be determined. Also, the preliminary evaluation of the average ship temperature model showed promise that it responds to the differing solar conditions. Future attempts at validation need to insure the accuracy of the ambient meteorological conditions. Onboard ground-truth radiometry measurements of the temperatures of different portions of the ship are also needed to aid in determining the accuracy of the adjusted average temperatures inferred from the AGA measurements. Finally, a controlled experiment with an airborne operational system needs to be conducted to determine the validity of the predicted detection ranges under varying meteorological conditions.

REFERENCES

[1] "Forward Looking Infrared Performance Function Program Performance Specification (PPS) for the Tactical Environmental Support System (TESS), Computer Sciences Corporation, NOSC TD 1000, December 1986.

[2] "FLIR Program Upgrade - Users Manual, STX Corporation, NEPRF Document 5W0513, UM-40, , August 1988.

[3] Cox, C. and W. Munk, "Measurement of Roughness of the Sea Surface from Photographs of the Sun's Glitter," Jour. Opt. Soc. of Am., 44, 838 (1954).

[4] Kneizys, F. X., E. P. Shettle, W. O. Gallery, J. H. Chetwynd, Jr., J. H. Abreu, J. E. A. Selby, S. A. Clough and R. W. Fenn, "Atmospheric Transmittance /Radiance: Computer Code LOWTRAN 6, Air Force Geophysics Laboratory Technical Report No. 83-0187, August 1983.

[5] Wollenweber, F. G., "Infrared Sea Radiance Modeling Using LOWTRAN 6," NOSC TD 1355, September 1988.

[6] Wollenweber, F. G., "Effects of Atmospheric Model Layering on LOWTRAN 6 Calculations of 8 to 12 μm Near Horizon Sky Radiances," NOSC TD 1193, January 1988.

[7] Hughes, H. G., "Evaluation of the LOWTRAN 6 Navy Maritime Aerosol Model Using 8 to 12 μm Sky Radiances," Opt. Eng., 26, 1155 (1987).

[8] Ostrowski, P. P. and D. M. Wilson, "A Simplified Computer Code for Predicting Ship Infrared Signatures," NSWC TR 84-540, November 1985.

[9] Lloyd, J. M., Thermal Imaging Systems, p. 28, Plenum Press, New York, 1975.

ACKNOWLEDGEMENTS

I wish to thank Fritz G. Wollenweber of the German Geophysical Office, FRG, who was responsible for incorporating the waveslope statistical model into LOWTRAN 6 and for many helpful discussions. Appreciation is also extended to Dr. Douglas R. Jensen who provided the airborne measurements of meteorological parameters, and to Merle R. Paulson who assisted in the AGA measurements. John Cook of the Naval Environmental Prediction Research Facility, Monterey, California was responsible for providing the FORTRAN version of SHIPSIG. Dr. Juergen H. Richter was responsible for providing the funding for this work effort. This work was sponsored by the Office of Navy Technology under Program Element 62435N, Project RM35G80, and Task N02C/02.

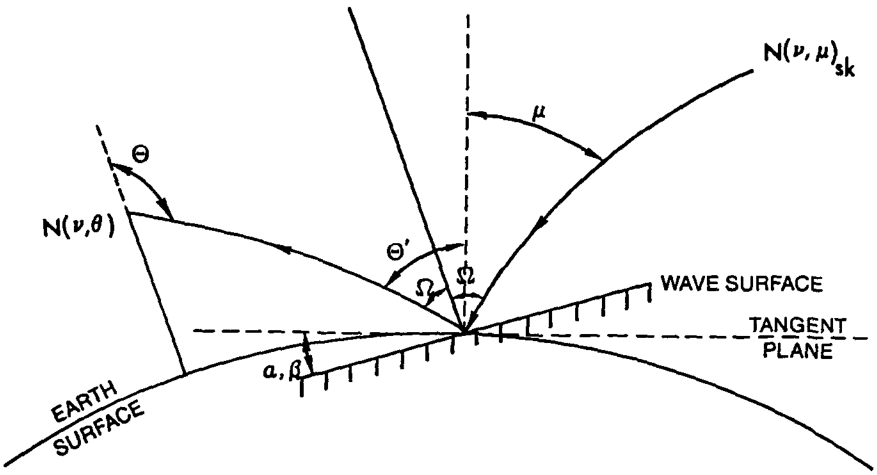


Figure 1. Reflection geometry from a wind-ruffled sea surface.

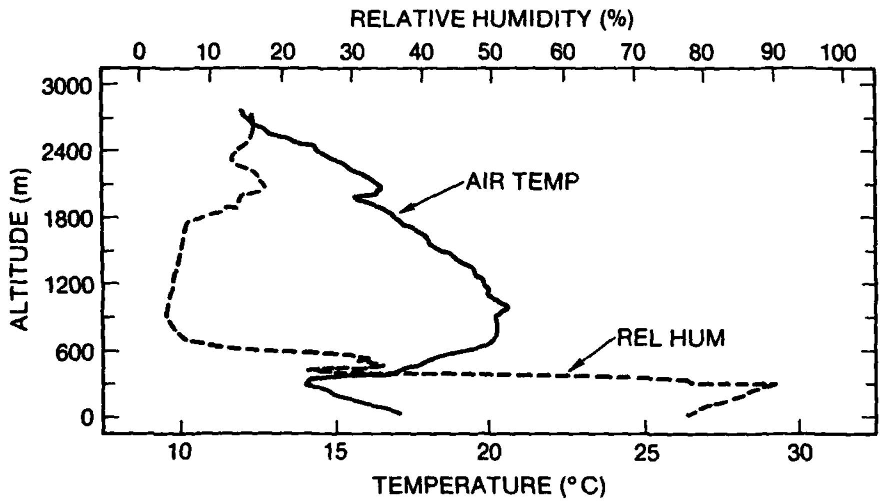


Figure 2. Profiles of air temperature and relative humidity measured with altitude on 9 June 1988 off the coast of San Diego, California.

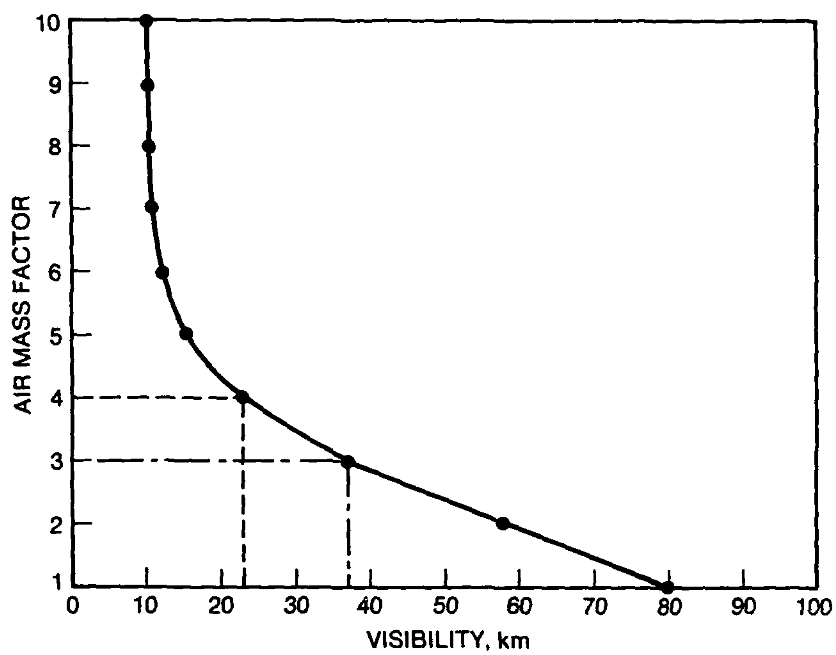


Figure 3. Locus of points of LOWTRAN 6 calculations with different combinations of air mass factors and visibilities which match the measured infrared horizon radiance in Figure 3.

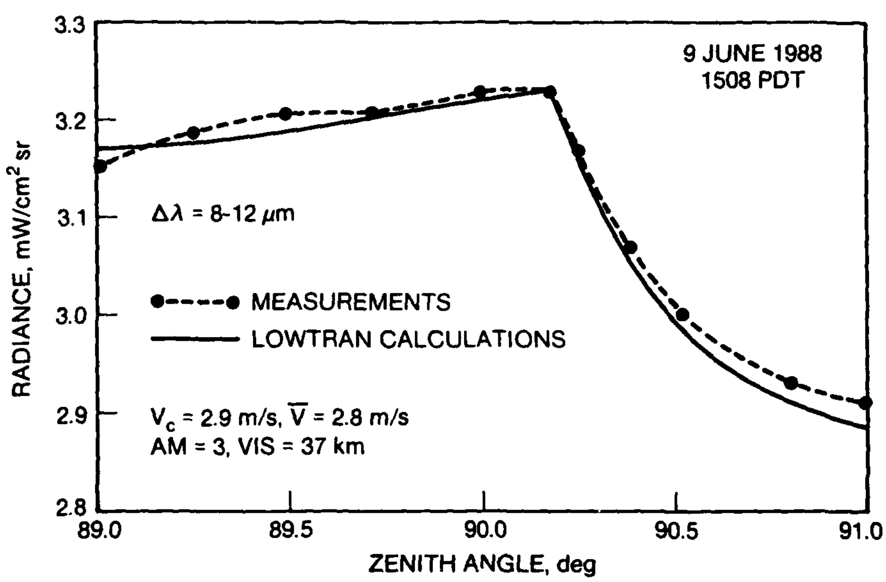


Figure 4. Comparison of measured and calculated infrared radiances for zenith angles above (θ < 90.17°) and below (θ > 90.17°) the horizon.

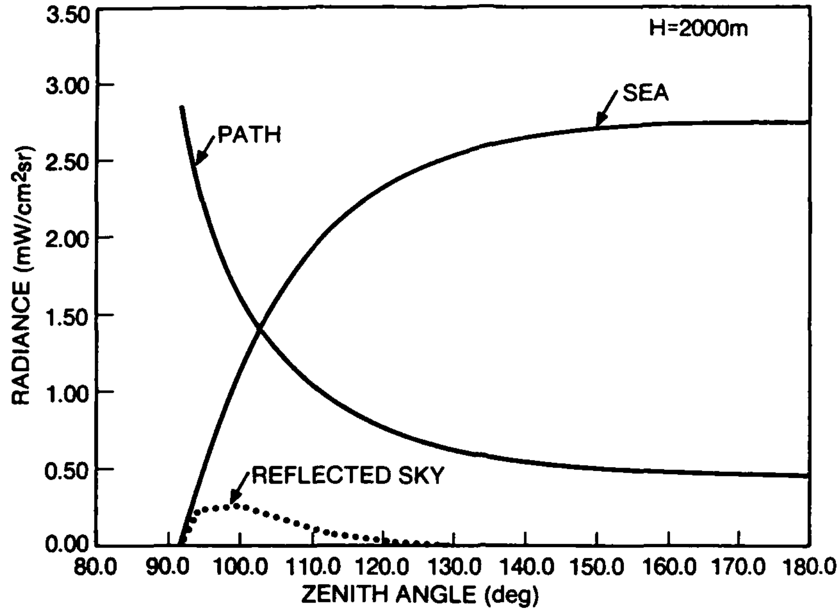


Figure 5. Individual contributions of the path, sea and reflected sky radiances to the total background radiance as a function of zenith angle for a sensor altitude of 2000 m.

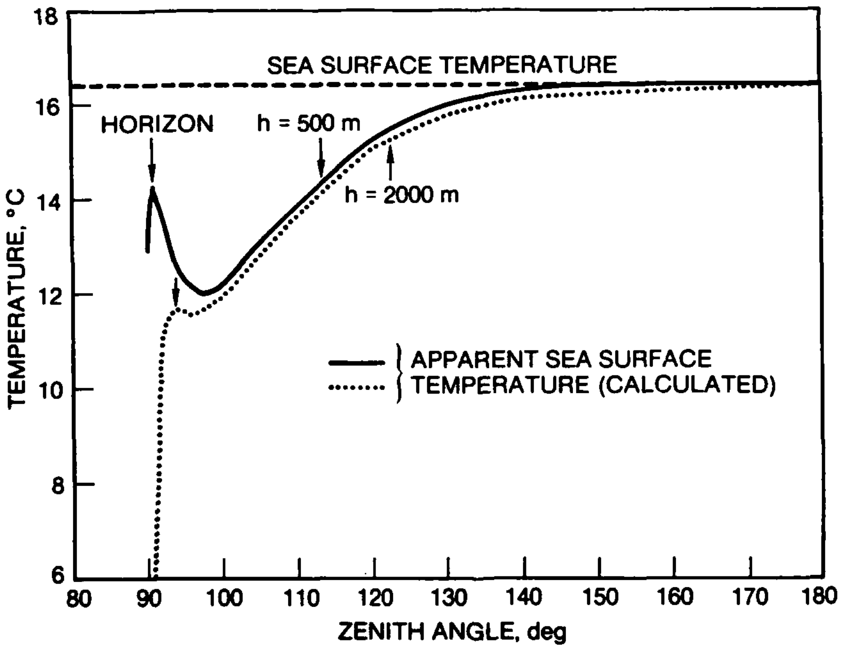


Figure 6. Total apparent blackbody temperature of the sea background versus zenith angle for sensor altitudes of 500 m and 2000 m.

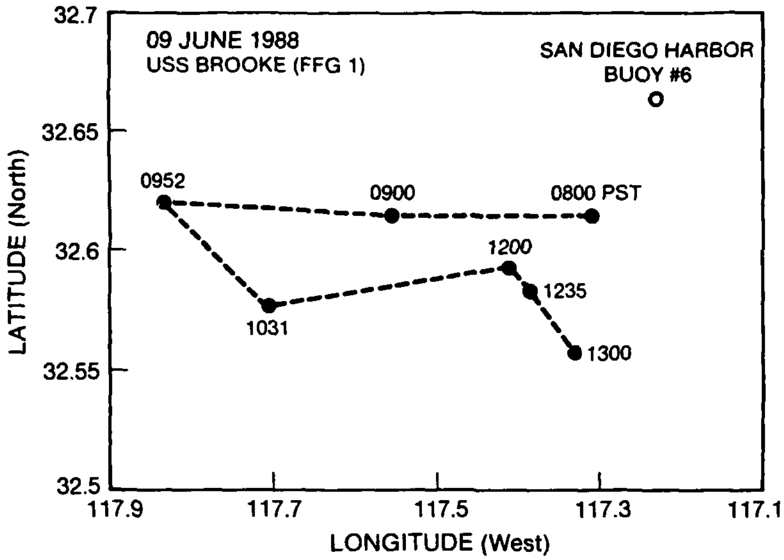


Figure 7. Course of USS BROOKE (FFG1) on 9 June 1988.

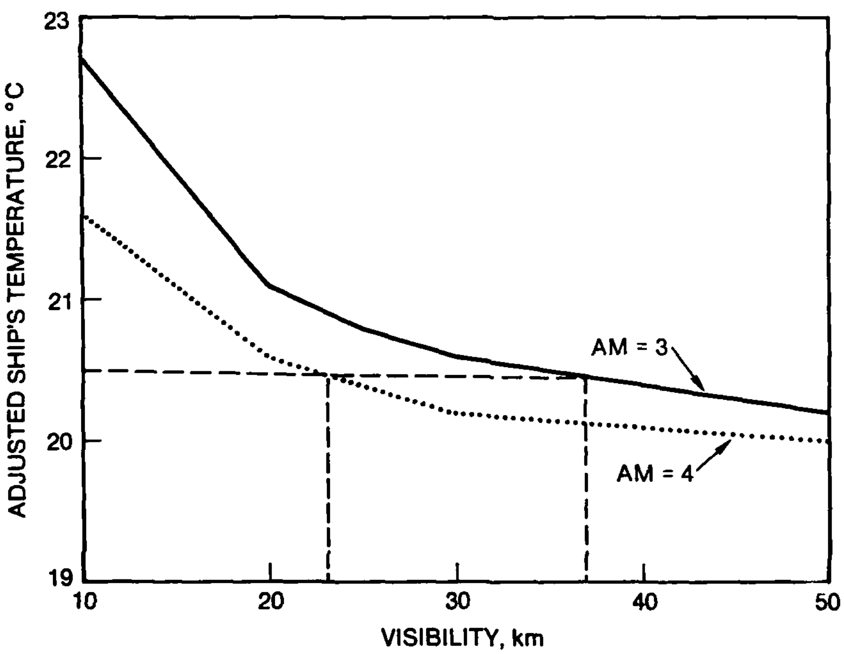


Figure 8. Ship's average temperature adjusted for atmospheric effects versus visibility for different air mass (AM) factors.

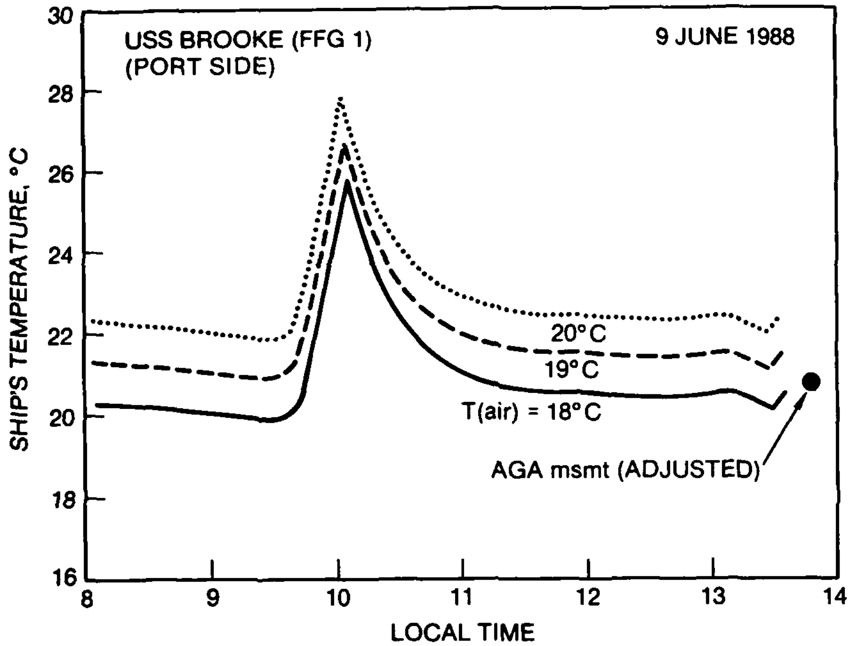


Figure 9. Comparison of the average temperatures of the port side of USS BROOKE (FFG1) (calculated using different ambient air temperatures) with the adjusted AGA measurements as the ship entered San Diego harbor.

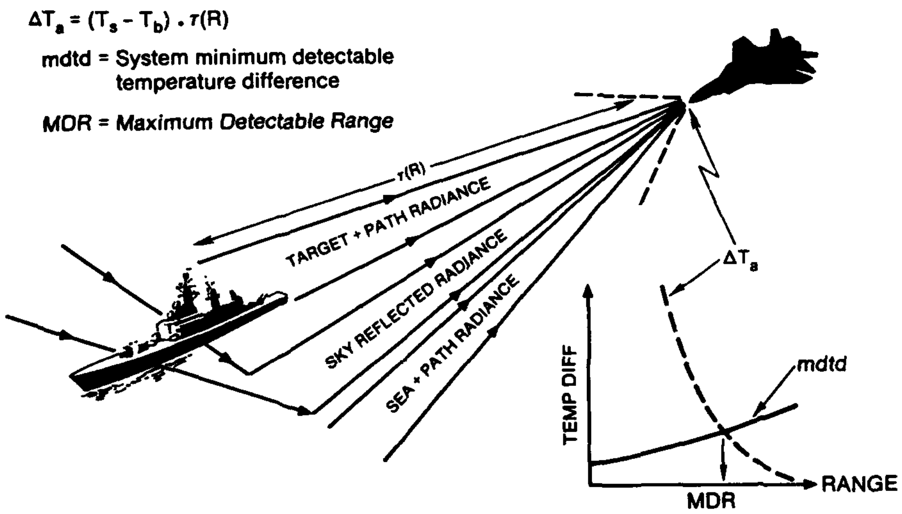


Figure 10. Illustration of the detection of a surface target by an airborne FLIR system.

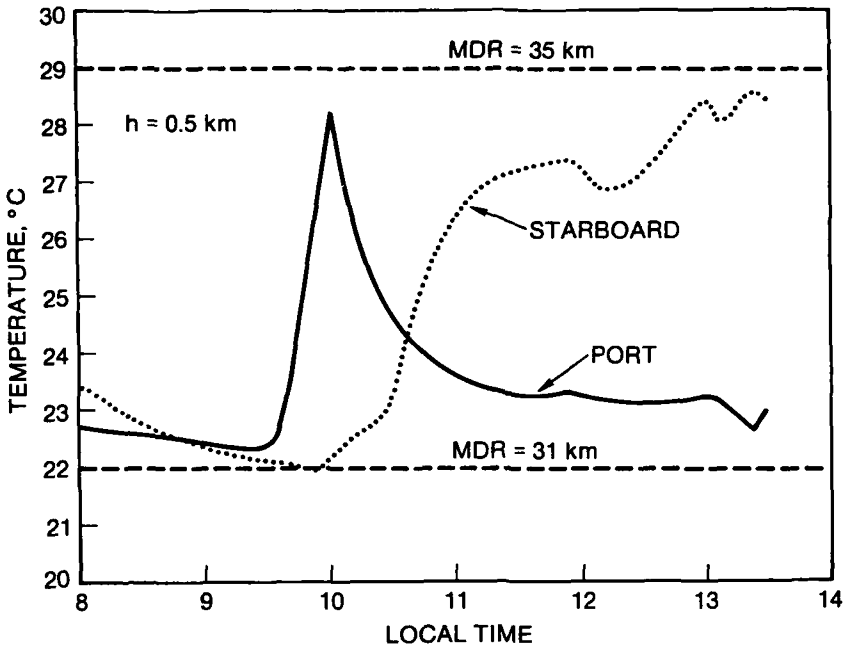


Figure 11. Calculated maximum detectable range (MDR) envelopes for USS BROOKE (FFG1) by an airborne FLIR at an altitude of 0.5 km.

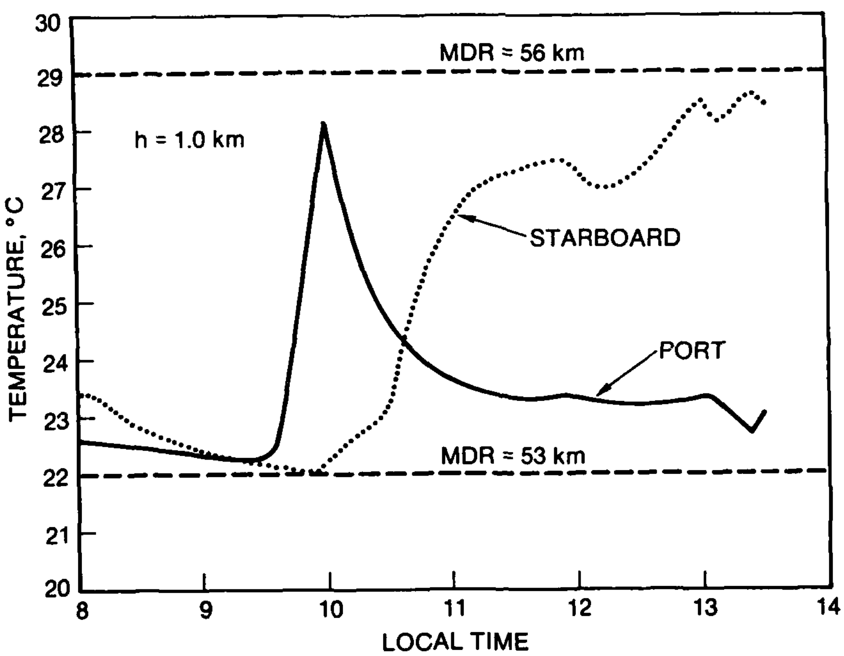


Figure 12. Calculated maximum detectable range (MDR) envelopes for USS BROOKE (FFG1) by an airborne FLIR at an altitude of 1.0 km.

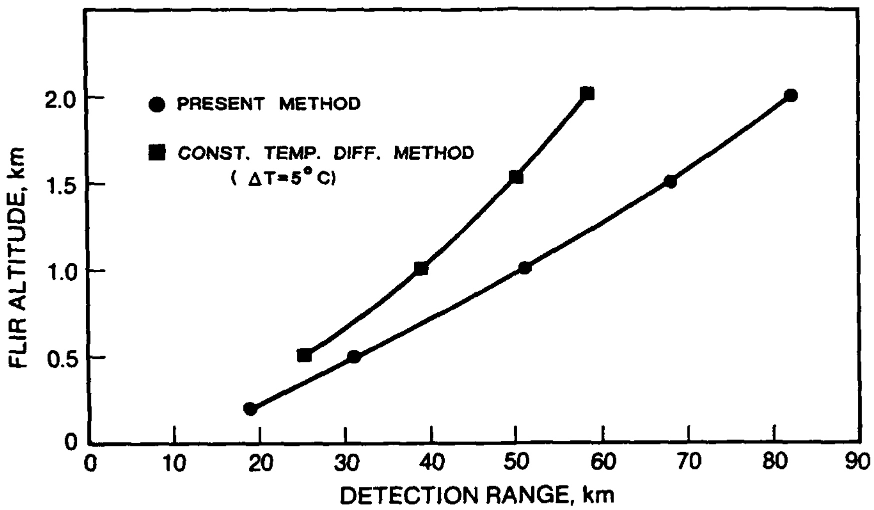


Figure 13. Comparison of the maximum detectable ranges (lower envelope) for USS BROOKE (FFG1) calculated with the present algorithm and those calculated assuming a constant temperature difference of 5°C between the ship and its background.

Hybrid PE-Ray-Mode Formulation of High Frequency Propagation in a Bilinear Tropospheric Surface Duct

T. Ishihara
Department of Electrical Engineering
National Defense Academy, Hashirimizu, Yokosuka, 239, Japan

and

L.B. Felsen
Department of Electrical Engineering/Computer Science
Polytechnic University, Route 110, Farmingdale, NY 11735 USA

Summary

Prediction of signal strengths due to high frequency sources in the presence of a tropospheric surface or elevated duct poses a problem of substantial complexity because the propagation environment is generally heterogeneous in height as well as in the lateral domain. A prediction algorithm should not only furnish adequate accuracy but also incorporate a parametrization in terms of physical observables so that the output can be properly interpreted. Relevant physical observables include ray fields, normal (trapped) and leaky mode fields, and beam-like (parabolic) propagators. These observables are employed here in self-consistent hybrid combination to model propagation in a laterally homogeneous surface duct with bilinear refractive index profile in height. A reference solution, which has been generated by normal and leaky mode summation, is interpreted qualitatively in terms of the corresponding ray field plot. Work is in progress on a quantitative reconstruction of the reference data via the hybrid algorithm.

I. Introduction

Analytical modeling of wave propagation in the earth's environment poses formidable problems because of the complexity of the propagation channel. The troposphere, which provides such a channel for electromagnetic waves, can generally be described by a permittivity $\varepsilon(r)$ that varies vertically as well as laterally with respect to the earth's surface. Of special interest are conditions where the vertical profile of $\varepsilon(r)$ is such as to permit wave ducting adjacent to the earth's surface or at moderate heights, with lateral variations occurring on a scale that is gradual compared with those in height. Because the problem scale is large at high frequencies and for long propagation distances, direct numerical modeling is either inefficient or not feasible. Therefore, an effective algorithm must be parametrized in terms of wave objects that can negotiate long range propagation. Possible choices include ray fields, mode fields, parabolic propagators, beams, etc., either for single species models or, more effectively, in hybrid combinations that seek to take advantage of the most favorable features of each. The analytical models must be self-consistent approximations of the intractable rigorously formulated problem. To ascertain their accuracy, they must be tested on tractable simpler canonical environments that allow generation of numerical reference solutions. Moreover, the test environments should be such as to deviate only weakly from the realistic environment, thereby validating analytic techniques of adiabatic adaptability of each wave object from the test problem to the realistic case without, or with only weak, coupling to other wave objects.

The present study seeks to apply these concepts to high frequency electromagnetic wave propagation in a tropospheric surface or elevated duct, initially without, and eventually with, lateral inhomogeneities. In the test environment, the permittivity is assumed to depend on height only, and the earth's influence is modeled by a constant surface impedance, thereby rendering the propagation problem separable in a spherical coordinate system. Because the (vertical electric dipole) source and observer locations of interest are at levels above the earth which are small compared to the earth's radius, it is possible to invoke approximations that map the spherical geometry into an equivalent rectangular geometry. A normalized field solution, usually referred to as the attenuation function, is then obtained in a spectral integral

form that serves as the starting point for developing desirable parametrizations. The considerations pertaining to this strategy have been well documented in the technical literature [1-3], and they are summarized in Sections II and III. In Section IV, detailed consideration is given to a bilinear permittivity profile which has also received much attention from other investigators [4]. After deriving the spectral integral solution for that special case (Section IV.A), alternative solutions are developed in the form of an expansion over normal (guided) modes propagating parallel to the earth's surface (Section IV.B), a ray field expansion (Section IV.C), and a hybrid ray-mode expansion that combines ray fields and mode fields self-consistently (Section IV.D).

II. Formulation

Adopting a spherical $\underline{r} = (r,\theta,\phi)$ coordinate system, we consider wave propagation due to a vertical electric dipole located near the earth's surface in a radially varying troposphere modeled by a permittivity $\epsilon(r)$. The electrical properties on the earth's surface at $r=a$ are assumed to be specified by a surface impedance Z_s. In view of the environmental dependence on r only, the radial dipole source excites fields that are azimuthally independent about its axis. If the polar axis of the spherical coordinate system is chosen to pass through the source at $r=r'$, the magnetic and electric fields are specified in that system by the components H_ϕ and E_r, E_θ, respectively (Fig. 1). All field components can be derived from a scalar potential $U(\underline{r}) = U(r,\theta)$ via the relations [1,2]

$$E_r = \frac{-1}{r\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial U}{\partial\theta} \tag{1}$$

$$E_\theta = \frac{1}{\epsilon(r)r}\frac{\partial}{\partial r}\left[\epsilon(r)r\frac{\partial U}{\partial\theta}\right] \tag{2}$$

$$H_\phi = i\omega\epsilon(r)\frac{\partial U}{\partial\theta} \tag{3}$$

with U(r) determined from the scalar wave equation

$$\frac{\partial}{\partial r}\frac{1}{\epsilon(r)}\frac{\partial}{\partial r}[\epsilon(r)(rU)] + \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta}U + k^2(r)rU = 0 \tag{4}$$

$$k^2(r) = \omega^2\epsilon(r)\mu_o = \omega^2\epsilon_o\mu_o\epsilon_s(r) = k_o^2\epsilon_s(r) \tag{4a}$$

In (4a), $\epsilon_s(r)$ is the permittivity normalized with respect to the permittivity ϵ_o in vacuum, and k_o is the wavenumber in vacuum. All field quantities have a suppressed time dependence $\exp(-i\omega t)$. The impedance boundary condition

$$E_\theta = -Z_s H_\phi \tag{5}$$

implies that

$$\frac{1}{\epsilon(r)r}\frac{\partial}{\partial r}[\epsilon(r)rU] = i\omega\epsilon(r)Z_s U \quad \text{on } r=a \tag{6}$$

At $|\underline{r}|\to\infty$, the potential must satisfy the radiation condition.

Because the radial locations of the source and observation points satisfy the inequalities $(r'/a) \ll 1$, $(r/a) \ll 1$, the exact equation in (4) can be approximated. By a sequence of scalings and redefinitions of coordinates, one may transform the field dependence from the spherical (r,θ) coordinates into an equivalent dependence in rectangular (y,x) coordinates [1,2]. The approximate source-excited transformed wave equation becomes

$$\left[\frac{\partial^2}{\partial y^2} + i\frac{\partial}{\partial x} + p(y)\right]\psi(x,y) = -\delta(x)\delta(y-y') \tag{7}$$

subject to the boundary condition

$$\frac{\partial}{\partial y}\psi + q\psi = 0 \quad \text{at } y=0 \tag{7a}$$

and a radiation condition at infinity. The following definitions have been utilized:

$$x = m\theta,\quad y = \frac{k}{m}(r\text{-}a),\quad y' = \frac{k}{m}(r'\text{-}a)\ , \tag{8a}$$

$$k = k(0),\quad m = \left(\frac{ka}{2}\right)^{1/3},\quad q = im[\epsilon_s(0)]^{1/2}Z_s/Z_o \tag{8b}$$

The "attenuation function" ψ is derived from the normalized potential

$$U/U_o = 2(\pi x)^{1/2}\, e^{i\pi/4}\, \psi\ , \tag{9}$$

where U_o is the potential for the same dipole located in the presence of a plane perfectly conducting earth. Because this normalization extracts the "strong" phase variation along $\theta \rightarrow (x/m)$, one obtains the simplified parabolic equation in (7). The equivalent permittivity function p(y) is given by

$$p(y) = y+\gamma(y),\quad \gamma(y) = m^2\left[\frac{\overline{\epsilon_s}(y)}{\overline{\epsilon_s}(0)} - 1\right],\quad \overline{\epsilon_s}(y) \equiv \epsilon_s(a+my/k) \tag{10}$$

This completes the formulation of the problem.

III. Spectral Integral Solution

The reduced equtION in (7) for the attenuation function is separable in the (x,y) coordinates. The most general form of the solution, from which various alternative representations can be derived, is built around a complex wavenumber spectrum. Introducing a complex spectral variable (separation parameter) t (this parameter should not be confused with the deleted time dependence exp(-iωt)), and recognizing that the $(\partial/\partial x)$ operator is algebraized by an exponential function exp(ixt) so that $(i\partial/\partial x) \rightarrow -t$, one obtains for the y-dependent spectrum the one-dimensional Green's function problem,

$$\left\{\frac{d^2}{dy^2} + [p(y)\text{-}t]\right\} g_y(y,y';t) = -\delta(y\text{-}y')\ , \tag{11}$$

which has the solution

$$g_y(y,y';t) = \frac{[f_2(y_<,t) + R_o f_1(y_<,t)]f_1(y_>,t)}{W[f_1,f_2]} \tag{12}$$

where the Wronskian W is given by

$$W[f_1,f_2] = f_1 f_2' - f_1' f_2\ ,\quad f' \equiv df/dt \tag{12a}$$

and the reflection coefficient R_o by

$$R_o = -\frac{f_2'(0,t) + qf_2(0,t)}{f_1'(0,t) + qf_1(0,t)} \tag{12b}$$

In (12), f_1 and f_2 are linearly independent solutions of the source-free equation in (11), which are employed in the combination shown to satisfy the boundary conditions at $y=0$ and $y\rightarrow\infty$, respectively, while $y_<$ and $y_>$ denote the smaller or larger values of y and y′, respectively. Spectral synthesis then yields the following solution,

$$\psi = \frac{1}{-2\pi i}\oint e^{ixt} g_y(y,y';t)dt \tag{13}$$

This result furnishes the starting point for the detailed studies that follow.

IV. Bilinear Permittivity Profile

A. Spectral integral

We shall consider for the equivalent permittivity p(y) in (10) the bilinear profile (Fig. 2)

$$\gamma(y) = (1+\mu^3)(y_i-y) \quad \text{for } 0<y\leq y_i \tag{14a}$$

$$\gamma(y) = 0 \quad \text{for } y_i\leq y \tag{14b}$$

where μ is a parameter to be determined from the profile expressed in the original (r,θ) coordinates. In (14), for subsequent convenience, the reference permittivity has been placed at $y=y_i$; accordingly, the reference permittivity $\overline{\epsilon_s}(0)$ in (10) should be replaced by $\overline{\epsilon_s}(y_i)$. If we write

$$\frac{\overline{\epsilon_s}(y)}{\overline{\epsilon_s}(y_i)} = 1-b_o\frac{m}{k}(y-y_i) \ , \quad y<y_i \tag{15}$$

then the profile parameters in (15) and (14) are related via

$$\mu = \left\{\frac{ab_o}{2} - 1\right\}^{1/3} \tag{16}$$

To reduce the equation in (11), with (14), to a standard form, we introduce the scaled coordinate and spectral variables ξ and τ,

$$\xi = \tau+\mu y\,, \quad \tau = [t-(1+\mu^3)y_i]/\mu^2 = \left\{t-\frac{ab_o y_i}{2}\right\}/\mu^2 \tag{17}$$

from which

$$\mu^2\xi_i = t-y_i \tag{17a}$$

where $\xi=\xi_i$ corresponds to $y=y_i$. Then the generic equation for the source-free solutions $f=f_{1,2}$ in (12) becomes the Airy equation

$$\frac{d^2}{d\xi^2}f-\xi f = 0 \quad \tau < \xi < \xi_i \tag{18}$$

Accordingly, the spectral integral in (13) may be written explicitly in terms of the Airy functions Ai, Bi,

$$\psi(x,y) = \frac{1}{2\mu i}\int_{-\infty}^{\infty}\frac{\overleftarrow{f}(y_<,t)\overrightarrow{f}(y_>,t)}{1-R_o(t)R_i(t)}\,e^{i\alpha t}\,dt \tag{19}$$

where

$$\overleftarrow{f}(y_<,t) = Ai(\tau+\mu y_<) + R_o Bi(\tau+\mu y_<) \tag{19a}$$

$$\overrightarrow{f}(y_>,t) = Bi(\tau+\mu y_>) + R_i Ai(\tau+\mu y_>) \tag{19b}$$

with the spectral reflection coefficients R_o and R_i that arise from the earth's surface at $y=0$ and the profile slope discontinuity at $y=y_i$, respectively,

$$R_o = -\frac{\mu Ai'(\tau)+qAi(\tau)}{\mu Bi'(\tau)+qBi(\tau)} \tag{19c}$$

$$R_i = -\frac{\mu Bi'(\tau+\mu y_i)+\dfrac{w_1'(t-y_i)}{w_1(t-y_i)}Bi(\tau+\mu y_i)}{\mu Ai'(\tau+\mu y_i)+\dfrac{w_1'(t-y_i)}{w_1(t-y_i)}Ai(\tau+\mu y_i)} \tag{19d}$$

with w_1 defined in (23). Alternative representations with different convergence properties will now be derived from the spectral continuum in (19).

B. Normal Mode Expansion

The integrand in (19) has pole singularities t_m at the zeros of the denominator (the index m here is not to be confused with the parameter m in (8b)),

$$1-R_o(t_m)R_i(t_m) = 0 \,, \quad m = 1,2,3,\cdots \qquad (20)$$

or explicitly

$$\frac{\mu Ai'(\tau_m)+qAi(\tau_m)}{\mu Bi'(\tau_m)+qBi(\tau_m)} = \frac{\mu w_1(t_m-y_i)Ai'(\tau_m+\mu y_i)+w_1'(t_m-y_i)Ai(\tau_m+\mu y_i)}{\mu w_1(t_m-y_i)Bi'(\tau_m+\mu y_i)+w_1'(t_m-y_i)Bi(\tau_m+\mu y_i)} \qquad (21)$$

By deforming the integration path in (19) around these poles in the upper half of the complex t-plane (the integrand converges at infinity so as to make this possible [2]) and invoking the residue theorem, one generates the residue series

$$\psi(x,y) = \frac{\pi}{\mu}\sum_{m=1}^{\infty} \frac{f_m(y')f_m(y)\exp(ixt_m)}{R_o(t_m)\frac{\partial}{\partial t}[R_o(t)R_i(t)]_{t_m}} \qquad (22)$$

where

$$f_m(y) = Ai(\tau_m+\mu y) + R_o(t_m)Bi(\tau_m+\mu y) \qquad (22a)$$

Each term in (22) represents a mode field $f_m(y)$ in the vertical cross section, which propagates along x with propagation coefficient t_m. The profile in (14) creates a surface duct between $y=0$ and $y=y_i$ because it is downward refracting in that height interval. For $y>y_i$, upward refraction takes place, with consequent leakage. The mode set t_m is therefore grouped into surface ducted (trapped) modes with small leakage ($\mathrm{Re}\,t_m>p(y_i)$), transitional modes ($\mathrm{Re}\,t_m\approx p(y_i)$), and leaky modes ($\mathrm{Re}\,t_m<p(y_i)$) (see Fig. 3). The estimates relating $\mathrm{Re}\,t_m$ to $p(y_i)$ are qualitative. Losses due to leakage out of the duct or due to a dissipative surface impedance Z_s are expressed by $\mathrm{Im}\,t_m$, which is positive; this controls the decay rate along x via the propagator $\exp(ixt_m)$ in (22), with larger values of $\mathrm{Im}\,t_m$ implying stronger damping.

C. Ray Expansion

The normal mode expansion in Section B is not always the most convenient for characterizing high frequency wave phenomena. An alternative is provided by a ray field expansion wherein local plane waves are tracked individually along direct and multiple reflected trajectories from source to observer [3,5]. Although the ray expansion can be phrased rigorously in terms of *generalized ray* spectral integrals, the principal utility of a ray field formulation rests in the asymptotic reduction of the generalized ray integrals, which yields the approximate fields that satisfy the rules of ray optics in inhomogeneous media. To generate the generalized ray series from the general result in (19), it is first necessary to decompose the oscillatory Airy functions Ai and Bi, which are natural descriptors of the (oscillatory) normal modes, into traveling wave form. Next, the resonant denominator $(1-R_oR_i)^{-1}$ in the integrand of (19) is removed by power series expansion, and the resulting integrand is rearranged into a sequence of combinations of traveling wave terms that form the generalized rays. Finally, asymptotic reduction produces the conventional ray fields.

Implementing the above scenario, we introduce the traveling wave Airy functions

$$w_{1,2} = Ai \mp iBi \qquad (23)$$

and thereby rewrite (19) in the form

$$\Psi(x,y_<,y_>) = \frac{-1}{4\mu}\int_C \frac{\overleftarrow{h}(y_<,t)\overrightarrow{h}(y_>,t)}{1-\overline{R}_o(t)\overline{R}_i(t)} e^{i\kappa t}dt \tag{24}$$

where

$$\overleftarrow{h}(y_<,t) = w_1(\tau+\mu y_<) + \overline{R}_o(t)w_2(\tau+\mu y_<) \tag{24a}$$

$$\overrightarrow{h}(y_>,t) = w_2(\tau+\mu y_>) + \overline{R}_i(t)w_1(\tau+\mu y_>) \tag{24b}$$

$$\overline{R}_o(t) = -\frac{\mu w_1'(\tau)+qw_1(\tau)}{\mu w_2'(\tau)+qw_2(\tau)} \quad , \quad w_{1,2}'(\tau) \equiv \frac{d}{d\tau}w_{1,2}(\tau) \tag{24c}$$

$$\overline{R}_i(t) = -\frac{\mu w_1(t-y_i)w_2'(\tau+\mu y_i)+w_1'(t-y_i)w_2(\tau+\mu y_i)}{\mu w_1(t-y_i)w_1'(\tau+\mu y_i)+w_1'(t-y_i)w_1(\tau+\mu y_i)} \tag{24d}$$

and

$$t-y_i = \mu^2(\tau+\mu y_i) \tag{24e}$$

Inserting the expansion

$$(1-\hat{R})^{-1} = \sum_{n=0}^{\infty}\hat{R}^n \quad , \quad \hat{R} = \overline{R}_o(t)\overline{R}_i(t) \; , \tag{25}$$

into (24), and interchanging the order of summation and integration, one may arrange the resulting series as follows,

$$\Psi = \sum_{j=1}^{4}\Psi^{(j)} \quad , \quad \Psi^{(j)} = \sum_{n=0}^{\infty}G_n^{(j)} \tag{26}$$

where $G_n^{(j)}$, $j=1...4$, represents four species of generalized ray integrals (Fig. 4)

$$G_n^{(1)} = \frac{-1}{4\mu}\int_C w_1(\tau+\mu y_<)w_2(\tau+\mu y_>)\hat{R}^n e^{i\kappa t}dt \tag{27a}$$

$$G_n^{(2)} = \frac{-1}{4\mu}\int_C w_2(\tau+\mu y_<)w_2(\tau+\mu y_>)\overline{R}_o\hat{R}^n e^{i\kappa t}dt \tag{27b}$$

$$G_n^{(3)} = \frac{-1}{4\mu}\int_C w_1(\tau+\mu_<)w_1(\tau+\mu y_>)\overline{R}_i\hat{R}^n e^{i\kappa t}dt \tag{27c}$$

$$G_n^{(4)} = \frac{-1}{4\mu}\int_C w_2(\tau+\mu y_<)w_1(\tau+\mu y_>)\overline{R}_o\overline{R}_i\hat{R}^n e^{i\kappa t}dt \tag{27d}$$

The physical content of the generalized ray fields may be understood by employing asymptotic approximations in the integrands, and then evaluating the integrals by the saddle point method. For large values of their arguments, the Airy functions $w_{1,2}$ in (23) can be approximated by [6]

$$w_{1,2}(-\nu) \sim \mp i\pi^{-1/2}\nu^{-1/4}\exp\left[\pm i\frac{2}{3}\nu^{3/2}\right] \pm i\pi/4, \quad , \quad Re\nu>0 \tag{28a}$$

$$w_{1,2}(\nu) \sim \mp i\pi^{-1/2}\nu^{-1/4}\exp\left[\frac{2}{3}\nu^{3/2}\right] \quad , \quad Re\nu>0 \tag{28b}$$

Propagating ray fields are established by progressing phase terms as in (28a). Because of the different arguments of the various $w_{1,2}$ functions in (27), one must identify distinct t-intervals where these arguments have positive or negative real parts. For the present purposes, to obtain propagating ray fields, it is required that a) $Re(\tau+\mu y_{\lessgtr})>0$, b) $Re\tau<0$; but one may have c) $Re(t-y_i)\gtrless 0$. As seen from (24d), the argument $(t-y_i)$ appears in the reflection coefficient due to the permittivity slope discontinuity at $y=y_i$. To distinguish the two cases, we append the subscript $\ell=1$ for $Re(t-y_i)>0$ and $\ell=2$ for $Re(t-y_i)<0$. The ray integrals $G_{\ell,n}^{(j)}$ pertaining to the spectral intervals C_ℓ defined by the constraints a), b) and $c_{1,2}$) are then reduced to the following forms (see Fig. 5 for typical trajectories):

$\ell = 1, j = 1$

Here, via (28b), $\overline{R}_i(t) \sim -i$, which implies that the upgoing waves in this spectral interval are evanescent at $y=y_i$. Virtual reflection takes place at a level $y<y_i$, and the value (-i) of the reflection coefficient identifies an encounter with a caustic. Then,

$$G^{(1)}_{\ell=1,n} \sim \frac{-1}{4\pi} \int_{C_1} (-i)^n \Gamma^n(t;p(0);q) \frac{\exp\{ix\Phi^{(1)}_{1,n}(t)\}}{\{p(y_<)-t\}^{1/4}\{p(y_>)-t\}^{1/4}} \tag{29}$$

where

$$\Gamma(t;p(0);q) = -\frac{q-i\sqrt{p(0)-t}}{q+i\sqrt{p(0)-t}} \tag{29a}$$

$$\Phi^{(1)}_{1,n}(t) = t+\frac{2}{3\mu^3 x}\left\{2n(p(0)-t)^{3/2} + (p(y_<)-t)^{3/2}-(p(y_>)-t)^{3/2}\right\} \tag{29b}$$

The forms for $j=2,3,4$ are similar. The saddle point $t^{(1)}_{1,n}$ of the integrand in (29) is located from the stationary phase condition

$$\frac{d}{dt}\Phi^{(1)}_{1,n}(t) = 0 \quad \text{at} \quad t^{(1)}_{1,n} \tag{30}$$

which yields the solution

$$x = \frac{1}{\mu^3}\left\{2n(p(0)-t)^{1/2}+(p(y_<)-t)^{1/2}-(p(y_>)-t)^{1/2}\right\}_{t=t^{(1)}_{1,n}} \tag{31}$$

Then by the conventional saddle point technique [7],

$$G^{(1)}_{1,n} \sim -\frac{1}{4}\sqrt{\frac{2}{\pi x Q^{(1)}_{1,n}}}\,(-i)^n\Gamma^n(t^{(1)}_{1,n};p(0);q)\,\frac{\exp\{ix\Phi^{(1)}_{1,n}(t^{(1)}_{1,n})+(i\pi/4)\operatorname{sgn}Q^{(1)}_{1,n}\}}{\{p(y_<)-t^{(1)}_{1,n}\}^{1/4}\{p(y_>)-t^{(1)}_{1,n}\}^{1/4}} \tag{32}$$

with

$$Q^{(1)}_{1,n} = \frac{d^2}{dt^2}\Phi^{(1)}_{1,n}(t)\bigg|_{t^{(1)}_{1,n}} = \frac{1}{\mu^3 x}\left\{2n\frac{1}{2\sqrt{p(0)-t}}+\frac{1}{2\sqrt{p(y_<)-t}}-\frac{1}{2\sqrt{p(y_>)-t}}\right\}_{t=t^{(1)}_{1,n}} \tag{32a}$$

The saddle point condition in (31) is the equation for a ray path reflected n-times at $y=0$, the term Γ^n in (32) incorporates the associated reflection coefficient at $y=0$, the exponential term in (32) describes the phase accumulation along that path, $(-i)^n$ accounts for n ray encounters with a caustic, and the remaining amplitude terms account for ray tube spreading.

$\ell = 2, j = 1$

Now, from (28a)

$$\overline{R}(t) \sim i\hat{\Gamma}(t;\mu;y_i)e^{-i\frac{4}{3\mu^3}(p(y_i)-t)^{3/2}} \tag{33}$$

where

$$\hat{\Gamma} \sim \frac{\mu+\{w_1'(\mu^2\xi_i)/w_1(\mu^2\xi_i)\}\,\{w_2(\xi_i)/w_2'(\xi_i)\}}{\mu+\{w_1'(\mu^2\xi_i)/w_1(\mu^2\xi_i)\}\,\{w_1(\xi_i)/w_1'(\xi_i)\}}\,,\quad \xi_i = \tau+\mu y_i \tag{33a}$$

Then

$$G^{(1)}_{2,n} \sim \frac{-1}{4\pi}\int_{C_2}\Gamma^n(t;p(0);q)\hat{\Gamma}^n(t;\mu;y_i)\,\frac{\exp\{ix\Phi^{(1)}_{2,n}(t)\}}{\{p(y_<)-t\}^{1/4}\{p(y_>)-t\}^{1/4}}\,dt \tag{34}$$

where

$$\Phi^{(1)}_{2,n}(t) = t+\frac{2}{3\mu^3 x}\left\{2n\left[(p(0)-t)^{3/2}-(p(y_i)-t)^{3/2}\right] + (p(y_<)-t)^{3/2}-(p(y_>)-t)^{3/2}\right\}, \tag{34a}$$

The saddle point condition yields

$$x = \frac{1}{\mu^3}\left\{2n\left[(p(0)-t)^{1/2}-(p(y_i)-t)^{1/2}\right] + (p(y_<)-t)^{1/2}-(p(y_>)-t)^{1/2}\right\} \tag{35}$$

and the asymptotic result

$$G^{(1)}_{2,n} \sim -\frac{1}{4}\sqrt{\frac{2}{\pi x Q^{(1)}_{2,n}}}\,\Gamma^n(t^{(1)}_{2,n};p(0);q)\hat{\Gamma}^n(t^{(1)}_{2,n};\mu;y_i)\,\frac{\exp\{ix\Phi^{(1)}_{2,n}(t^{(1)}_{2,n})+(i\pi/4)\operatorname{sgn}Q^{(1)}_{2,n}\}}{\{p(y_<)-t^{(1)}_{2,n}\}^{1/4}\{p(y_>)-t^{(1)}_{2,n}\}^{1/4}}, \tag{36}$$

$$Q^{(1)}_{2,n} = \frac{1}{\mu^3 x}\left\{2n\left[\frac{1}{2\sqrt{p(0)-t}} - \frac{1}{2\sqrt{p(y_i)-t}}\right] + \frac{1}{2\sqrt{p(y_<)-t}} - \frac{1}{2\sqrt{p(y_>)-t}}\right\}_{t=t^{(1)}_{2,n}} \tag{36a}$$

Similar results can be derived for the other species $j=2,3,4$, $\ell=1,2$.

The forms of the solution for $\ell = 1,2$ and $j=1$ in (32) and (36) are generic also for the other (ℓ,j) combinations. Thus

$$G^{(j)}_{\ell,n} \sim -\frac{1}{4}\sqrt{\frac{2}{\pi x Q^{(j)}_{\ell,n}}}\,\Gamma^n(t^{(j)}_{\ell,n};p(0);q)\hat{\Gamma}^n(t^{(j)}_{\ell,n};\mu;y_i)\,\frac{\exp[ix\Phi^{(j)}_{\ell,n}(t^{(j)}_{\ell,n})+(i\pi/4)\operatorname{sgn}Q^{(j)}_{\ell,n}]}{\sqrt{\Omega(y_<,t^{(j)}_{\ell,n})\Omega(y_>,t^{(j)}_{\ell,n})}} \tag{37}$$

$$\Omega(y,t) \equiv [p(y)-t]^{1/2} \quad . \tag{37a}$$

Here, Γ is defined in (29a), and $\hat{\Gamma}$ in (33a), for $\ell=2$, whereas $\hat{\Gamma}=-i$ for $\ell=1$. Moreover,

$$\begin{aligned}\Phi^{(j)}_{\ell,n} = t^{(j)}_{\ell,n} + \frac{2}{3\mu^3 x}\Big\{&(2n+\alpha_1)\left[\alpha_2\Omega^3(0,t)-\alpha_3\Omega^3(y_i,t)\right]+\alpha_2'\Omega^3(0,t)\\ &-\alpha_3'\Omega^3(y_i,t)+\alpha_4\Omega^3(y_<,t)+\alpha_5\Omega^3(y_>,t)\Big\}_{t^{(j)}_{\ell,n}}\end{aligned} \tag{37b}$$

$$\begin{aligned}Q^{(j)}_{\ell,n} = \frac{1}{2\mu^3 x}\Big\{&(2n+\alpha_1)\left[\alpha_2\Omega^{-1}(0,t)-\alpha_3\Omega^{-1}(y_i,t)\right]+\alpha_2'\Omega^{-1}(0,t)-\alpha_3'\Omega^{-1}(y_i,t)\\ &+\alpha_4\Omega^{-1}(y_<,t)+\alpha_5\Omega^{-1}(y_>,t)\Big\}_{t^{(j)}_{\ell,n}}\end{aligned} \tag{37c}$$

and the ray equation (saddle point condition) is

$$x = \frac{1}{\mu^3}\Big\{(2n+\alpha_1)\left[\alpha_2\Omega(0,t)-\alpha_3\Omega(y_i,t)\right]+\alpha_2'\Omega(0,t)-\alpha_3'\Omega(y_i,t) \tag{37d}$$

$$+\ \alpha_4\Omega(y_<,t)+\alpha_5\Omega(y_>,t)\Big\}_{t^{(j)}_{\ell,n}} \tag{37d}$$

Then the following definitions of the constants α_i, $i=1$ to 5, apply:

$$\ell=1,\ j=2:\ \alpha_1=\alpha_2=-\alpha_4=-\alpha_5=1,\ \alpha_2'=\alpha_3'=\alpha_3=0 \tag{38a}$$

$$\ell=2,\ j=2:\ \alpha_1=\alpha_3'=0,\ \alpha_2'=2,\ \alpha_3=-\alpha_4=-\alpha_5=1 \tag{38b}$$

$$\ell=1,\ j=3:\ \alpha_1=\alpha_3=\alpha_2'=\alpha_3'=0,\ \alpha_2=\alpha_4=\alpha_5=1 \tag{38c}$$

$$\ell=2,\ j=3:\ \alpha_1=\alpha_2'=0,\ \alpha_2=\alpha_4=\alpha_5=1,\ \alpha_3'=2 \tag{38d}$$

$$\ell=1,\ j=4:\ \alpha_1=2,\ \alpha_2=-\alpha_4=\alpha_5=1,\ \alpha_3=\alpha_2'=\alpha_3'=0 \tag{38e}$$

$$\ell=2,\ j=4:\ \alpha_1=2,\ \alpha_2=\alpha_3=-\alpha_4=\alpha_5=1,\ \alpha_2'=\alpha_3'=0 \tag{38f}$$

Typical ray paths are identified in Fig. 5. All of the asymptotic ray field results above are based on the validity of the asymptotic approximations in (28) for the integrand, and also on the validity of the isolated saddle point evaluation. This excludes observation points near cautics, and also near glancing rays that are tangent to the refractive index slope discontinuity (duct boundary) at $y=y_i$ (Fig. 5).

D. Hybrid Ray-Mode Expansion

The most general format involving ray fields and mode fields is obtained by combining these self-consistently in a hybrid ray-mode expansion [5]. To generate the hybrid representation, we first deform the integration path C in (19) into the two contours $C_\ell, \ell=1,2$, which surround, respectively, the trapped modes with $\mathrm{Re}\, t_m > p(y_i)$ and the leaky modes with $\mathrm{Re}\, t_m < p(y_i)$; these contours (Fig. 6) establish the rigorous complex extension of the real spectrum intervals associated with $\ell=1$ and $\ell=2$ in Section IV.C. Next, we expand the resonant denominator not into an infinite (ray) series as in (25) but into a truncated series:

$$\frac{1}{1-\hat{R}} = \frac{1}{1-\hat{R}} - \frac{1}{2}\frac{1+\hat{R}}{1-\hat{R}}\hat{R}^{N_2} - \frac{1}{2}\hat{R}^{N_2} + \sum_{n=N_2}^{N_1-1}\hat{R}^n + \frac{1}{2}\hat{R}^{N_1} + \frac{1}{2}\frac{1+\hat{R}}{1-\hat{R}}\hat{R}^{N_1} \tag{39}$$

Then decomposing the attentuation function corresponding to C_1 and C_2 as in (26), one obtains

$$\psi = \sum_{j=1}^{4}\sum_{\ell=1}^{2}\psi_\ell^{(j)} \tag{40a}$$

where

$$\psi_\ell^{(j)} = \sum_{n=N_2}^{N_1-1} G_{\ell,n}^{(j)} + \frac{1}{2}G_{\ell,N_1}^{(j)} - \frac{1}{2}G_{\ell,N_2}^{(j)} + R_{\ell,N_1}^{(j)} + R_{\ell,N_2}^{(j)} \tag{40b}$$

Here, $G_{\ell,n}^{(j)}$ are the generalized ray integrals associated with C_ℓ, $\ell=1,2$, which have the integrands shown in (27). The remainder integrals $R_{\ell,N_\ell}^{(j)}$ have for each wave species j=1 to 4 the same integrand as the corresponding ray integral $G_{\ell,N_\ell}^{(j)}$, but contain in the integrand the following additional factors:

$$R_{\ell,N_1}^{(j)}:\ M,\ M \equiv \frac{1}{2}(1+\hat{R})(1-\hat{R})^{-1}\ ,\ j=1 \text{ to } 4 \tag{41a}$$

$$R_{\ell,N_2}^{(j)}:\ [\hat{R}^{N_2}(1-\hat{R})]^{-1} - M\ ,\ j=1 \text{ to } 4 \tag{41b}$$

Asymptotic reduction of the ray integrals proceeds as in Section IV.C. The remainder integrals can be reduced by deforming the integration paths $C_{1,2}$ into steepest descent paths $SDP^{(j)}$ through the saddle point $t_{\ell,N_\ell}^{(j)}$ of the corresponding ray integral. Because each integrand contains the resonant denominator $(1-\hat{R})^{-1}$, some of the modal pole singularities t_m are encountered during the deformation (Fig. 7) and contribute residue contributions that furnish partial modes ("partial" implies that the decomposed traveling wave spectra represented by w_1 and w_2 in the integrands contribute individually, instead of the full spectra expressed by the Ai, Bi functions in (22)). For $\ell=j=1$ and N_1, one obtains from deformation of C_1,

$$R_{1,N_1}^{(1)} = \sum_{m=1}^{M_{1,N_1}^{(1)}} G_m^{(1)} + R_{M_{1,N_1}^{(1)}} \tag{42}$$

where

$$G_m^{(1)} = \frac{\pi i}{2\mu}\,\frac{w_1(\tau_m+\mu y_<)w_2(\tau_m+\mu y_>)}{\frac{\partial}{\partial t}\bar{R}_o(t)\bar{R}_i(t)\,|_{t=t_m}}\,e^{ixt_m} \tag{42a}$$

is the partial mode field, and $R_{M_1^{(1)},N_1}$ is a new remainder integral having the same integrand as $R_{1,N_1}^{(1)}$ but extending along $SDP_{1,N_1}^{(1)}$. For saddle point $t_{1,N_1}^{(1)}$ separated from the modal pole $t_{M^{(1)}}$, the new remainder integral can be evaluated by the saddle point method to yield

$$R_{M_{1,N_1}^{(1)}} \sim \frac{1}{2}G_{1,N_1}^{(1)}\cdot\underline{\Gamma} \tag{43}$$

where

$$\underline{\Gamma} = \left.\frac{1+\bar{R}_o(t)\bar{R}_i(t)}{1-\bar{R}_o(t)\bar{R}_i(t)}\right|_{t=t_{1,N_1}^{(1)}} \sim \left.\frac{e^{-i\{\frac{2}{3}(-\tau)^{3/2}+\pi/4\}} - e^{i\{\frac{2}{3}(-\tau)^{3/2}+\pi/4\}}\Gamma}{e^{-i\{\frac{2}{3}(-\tau)^{3/2}+\pi/4\}} + e^{i\{\frac{2}{3}(-\tau)^{3/2}+\pi/4\}}\Gamma}\right|_{t_{1,N_1}^{(1)}} \tag{43a}$$

and

$$\Gamma = \Gamma(t;p(0);q) = -\frac{q-i\sqrt{p(0)-t}}{q+i\sqrt{p(0)-t}} \tag{43b}$$

For a perfectly conducting earth ($q=\infty$), (43) reduces to

$$R_{M^{(1)}_{1,N_1}} \sim \frac{i}{2} G^{(1)}_{1,N_1} \cot\left\{\frac{2}{3\mu^3}\left(p(0)-t^{(1)}_{1,N_1}\right)^{3/2}+\pi/4\right\} \tag{44}$$

The contribution in (43) or (44) is usually small. When the saddle point is near the modal pole, the new remainder integral cannot be neglected and must be evaluated by uniform asymptotics in terms of incomplete error functions [7].

For $\ell = j = 1$, and N_2, the remainder integral is split into two parts (see (41b)). For the first term in (41b), the integration path C_1 is deformed around the enclosed pole singularities to generate a sum of modes as in (42a). For the second term in (41b), C_1 is deformed into $SDP^{(1)}_{1,N_2}$. This yields

$$R^{(1)}_{1,N_2} = \sum_{m=q}^{M} G^{(1)}_m - R_q \quad , \quad q \equiv M^{(1)}_{1,N_2} \tag{45}$$

The integral R_q is the same as the one for $R^{(1)}_{1,N_1}$ arising from the second term of (41b), except that the integration path is $SDP^{(1)}_{1,N_2}$. Therefore, the evaluation of the integral can proceed as above. The same scenario can be repeated for all of the other integrals $R^{(j)}_{\ell,N_\ell}$. (Also, the PE (parabolic equation) algorithm can be used to replace certain mode groups [5]. These aspects will be dealt with in detail in a future publication.) When all of these results are combined, one obtains the following most general hybrid ray-mode representation of the attenuation function:

$$\psi = \sum_{j=1}^{4}\sum_{\ell=1}^{2}\left\{\sum_{n=N_2}^{N_1-1} G^{(j)}_{\ell,n} + \frac{1}{2}G^{(j)}_{\ell,N_1} - \frac{1}{2}G^{(j)}_{\ell,N_2}\right\}$$

$$+ \sum_{j=1}^{4}\sum_{m=1}^{M^{(j)}_{1,N_1}} G^{(j)}_m + \sum_{j=1}^{4}\sum_{m=M^{(j)}_{1,N_3}}^{M^{(j)}_{2,N_3}} G^{(j)}_m + \sum_{j=1}^{4}\sum_{m=M^{(j)}_{2,N_1}}^{\infty} G^{(j)}_m$$

$$+ \sum_{j=1}^{4}\sum_{\ell=1}^{2}\left(R_{M^{(j)}_{\ell,N_1}} - R_{M^{(j)}_{\ell,N_3}}\right) \tag{46}$$

The expansion in (46) is *exact* if the generalized ray integrals and all remainder integrals are kept intact, but useful results are obtained by employing the asymptotic reductions discussed earlier. We have concentrated here on formulations that utilize asymptotic ray theory with added corrections near isolated caustics. When these approximations fail, the illegitimate ray fields are to be replaced by modes. Such failures occur around the glancing ray transition (with energy splitting) at the upper duct boundary, and also when source and observer are located near the bottom of the duct, thereby generating trapped rays with many reflections that give rise to an accumulation of caustics. These phenomena are schematized in Fig. 5, and more specifically in Fig. 8. The mode cluster, which eliminates the illegitimate near-glancing rays at the upper boundary of the duct, remains to be established. The few modes that can account for the illegitimate multiple rays near the bottom may in turn be replaced by a parabolic equation (PE) propagator [5]. Clarification of these issues is now in progress. Also under investigation are the conditions that allow the partial modes $G^{(j)}_m$ in (46) to be combined into full modes G_m, and (or) allow the remainder integrals to be simplified or neglected.

V. Numerical Results

Extensive numerical calculations have been initiated to test the validity of the various hybrid combinations described above. The reference solutions for comparison are being generated from the complete mode expansion in (22). Results obtained so far for one of the test cases are shown in Fig. 9, and may be interpreted by recourse to the ray diagram in Fig. 8. The duct parameters (cf. (10) and (14), and Fig. 2) are $y_i = 22.34$, $p(0) = 93.51$, $p(y_i) = 22.34$, $b_o = 1.313972 \times 10^{-6}$, q (impedance) = 0, and f (frequency) = 9 GHz. The source is located on the lower boundary, which is assumed to be perfectly reflecting. Due to constructive and destructive interference among the many propagating modes (107 modes have been included here), the resulting field is far from mode-like. Instead, the physical propagation mechanism is well explained in terms of ray fields, as schematized in Fig. 8.

The field magnitude plots vs. range in Fig. 9 refer to observer heights of 80 m (Fig. 9a), 140 m (Fig. 9b) and 180 m (Fig. 9c), respectively; the duct extends up to 200 m. The range points identified on these plots correspond to those in Fig. 8, where they are seen to locate distinctive regions pertaining to the ray diagram. For the 80 m observer height, the range interval $A'A_1$ in Fig. 8 is reached only by a single (the direct) ray. The ray field magnitude in Fig. 9a decays monotonically there, in accord with the single ray interpretation. Between A_1 and A_2, the observer receives two ray contributions (each having one bottom reflection), thereby interpreting the almost periodic oscillations in Fig. 9a that accompany this interference mechanism. The A_2A_3 interval is again reached only by a single ray (with one bottom reflection) so that the field decay resembles that in the $A'A_1$ interval. The irregular interference pattern between A_3 and A_4 in Fig. 9a is attributed to the presence of 3 rays in Fig. 8, whereas the more regular pattern in the interval A_4A_5 is again due to 2 rays. When the observer moves to an elevation of 140 m, he encounters ray shadow zones in range intervals B_1B_2, B_4B_5, B_7B_8 (Fig. 8). Accordingly, the field there is very weak (see Fig. 9b), and is, in fact, not predicted by the infinitely dark shadow (zero field) of simple ray theory; however, uniform ray theory [7] can account for penetration of the shadow zone. The field behavior in the illuminated regions follows the format discussed for the 80 m elevation, in view of the following ray configuration here: $B'B_1$ - direct ray only; B_2B_3 - 2 rays (B_2 lies on a ray caustic and B_3 on a limiting ray); B_3B_4 - 1 ray; B_5B_6 - 2 rays; B_6B_7 - 1 ray. These phenomena are sharpened at the 180 m elevation, which approaches the upper duct boundary. The simple field magnitude profile in Fig. 9c follows from the ray configuration (Fig. 8): $C'C_1$ - direct ray only; C_1C_2 - shadow; C_2C_3 - 1 ray; C_3C_4 - shadow; C_4C_5 - 1 ray.

It has therefore been confirmed that the qualitative interpretation of the reference field data in Fig. 9, although generated by modal summation, is actually linked in the most direct and straightforward manner to the ray fields. The quantitative prediction of the fields in Fig. 9 by the ray model evidently requires refinements, especially in and near the critical regions identified earlier. This is where the hybrid form comes into play. An investigation of this aspect is in progress.

VI. Summary

We have established the machinery for systematic investigation of high frequency electromagnetic propagation in a troposphere whose height refractive index profile gives rise to a surface duct. A bilinear index profile has been chosen for a test environment because accurate numerical reference solutions, based on an exact modal sum, can be generated in this case. Our goal has been to parametrize the field in terms of ray field constituents, which have cogent physical content and can be computed with relative ease, and to compensate the simple asymptotic ray algorithm with other field constituents in those critical regions where ray theory fails. This has led to a self-consistent hybrid format, which is given in its most general form in (46). The format is rigorous and therefore allows systematic study of simplifying assumptions that reduce its complexity. Such a study is now in progress. It is hoped that there will emerge an algorithm capable of making reliable predictions in those observational regions that are not treatable by simple ray theory alone, without adding excessive computational complexity.

If our endeavor succeeds for the test environment, it can be generalized to other height profiles, including those with weak lateral and longitudinal variations. The generalization will be based on the localization and adaptability of high frequency wave phenomena. Ray fields are readily defined in a three-dimensionally varying environment, and local modes can do the same for modal constituents. Extensive numerical experiments, and comparisons with other algorithms (for example, PE) as well as reliable experimental data, will have to be carried out to establish the limits of accuracy of the hybrid model.

References

[1] J.R. Wait, *Electromagnetic Waves in Stratified Media*, Pergamon Press, New York, pp. 107-137, pp. 341-363 (1962).

[2] V.A. Fock, *Electromagnetic Diffraction and Propagation Problems*, Pergamon Press, New York, pp. 254-275 (1965).

[3] A.V. Kukushkin and V.G. Sinitsin, "Rays and Modes in a Nonuniform Troposphere," Radio Science, Vol. 18, No. 4, pp. 573-587 (1983).

[4] H.V. Hitney, R.A. Pappert and C.P. Hattan, "Evaporation Duct Influences on Beyond-the-Horizon High Altitude Signals," Radio Science, Vol. 13, No. 4, pp. 669-675 (1978). This paper has a general overview.

[5] T. Ishihara and L.B. Felsen, "Hybrid (Ray)-(Parabolic Equation) Analysis of Propagation in Ocean Acoustic Guiding Environments," J. Acoust. Soc. Am., vol. 83, No. 3, pp. 950-960 (1988).

[6] A. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions*, Dover Publications, pp. 437-478 (1972).

[7] L.B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves*, Prentice-Hall, Inc., New Jersey, pp. 399-406 and Chapter 5 (1973).

Acknowledgement

This work has been supported by the Ocean and Atmospheric Sciences Division, Naval Ocean Systems Center, San Diego, California.

Figure Captions

Fig. 1 Spherical earth (radius $r=a$) with vertical electric dipole source at Q. (r,θ,ϕ) are spherical coordinates, and the excited field components are E_r, E_θ, H_ϕ.

Fig. 2 Bilinear permittivity profile corresponding to (10) and (14).

Fig. 3 Qualitative placement of singularities (for low loss surface conditions) and integration path for (19) in complex t-plane. Trapped and leaky mode ranges correspond roughly to $\mathrm{Re}\, t_m > p(y_i)$ and $\mathrm{Re}\, t_m < p(y_i)$, respectively, where y_i locates the upper boundary of the duct. m = mode index.

Fig. 4 Ray species $j=1 \cdots 4$ for categories $\ell=1$ (trapped inside duct) and $\ell=2$ (leaking out of duct). For arbitrary reflection index n, the ray species are organized according to their directions of departure from Q and arrival at P.

Fig. 5 Typical ray trajectories categorized as in Fig. 4. Trapped rays ——; leaky rays -·-·-; glancing rays ---.

Fig. 6 Deformed integration contours preliminary to hybrid ray-mode expansion. C_1 and C_2 enclose the trapped and leaky modes, respectively.

Fig. 7 Hybrid ray-mode expansion by deformation of integration contours C_1 and C_2 into steepest descent paths $\mathrm{SDP}^{(j)}_{\ell,N_\ell}$ through the saddle points $t^{(j)}_{\ell,N_\ell}$ of the remainder field integrands. Mode contributions arise from poles crossed during the deformation.

Fig. 8. Ray plot drawn to scale for bilinear surface duct with parameters h = 200m (y_i = 22.34), p(0) = 93.51, p(y_i) = 22.34, b_o = 1.313972x10^{-6}, q = 0, and f = 9GHz. The source is located on the perfectly reflecting bottom boundary. Observation heights at 80m (A´A´´), 140 m (B´B´´) and 180 m (C´C´´) correspond to the numerical data in Fig. 9. The domains illuminated by various ray species are bounded by their caustics and by limiting rays (which graze the upper duct boundary).

Fig. 9. Field magnitude vs. horizontal range for the three elevations identified in the duct of Fig. 8.

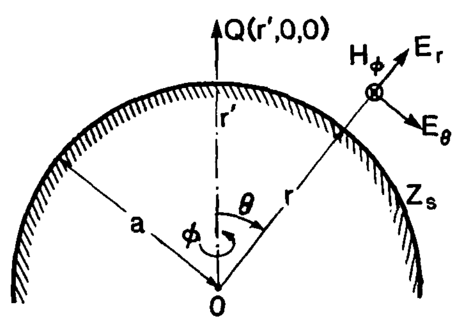


Fig. 1

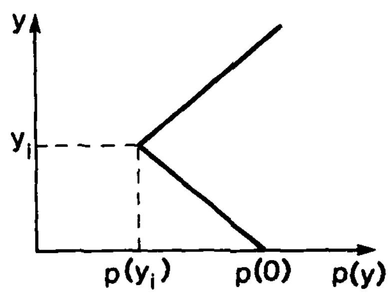


Fig. 2

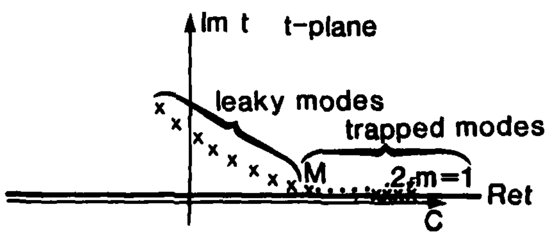


Fig. 3

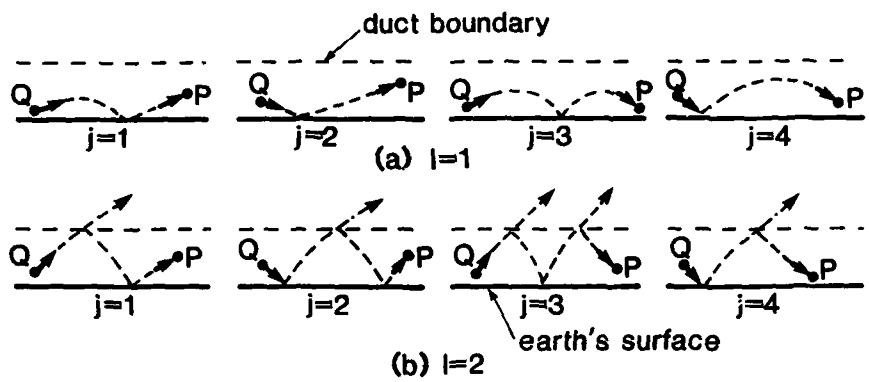


Fig. 4

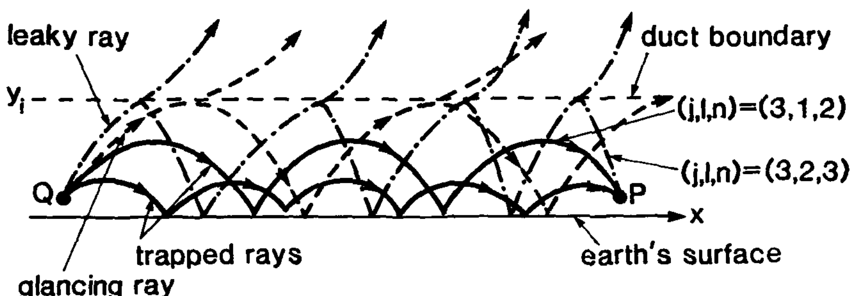


Fig. 5

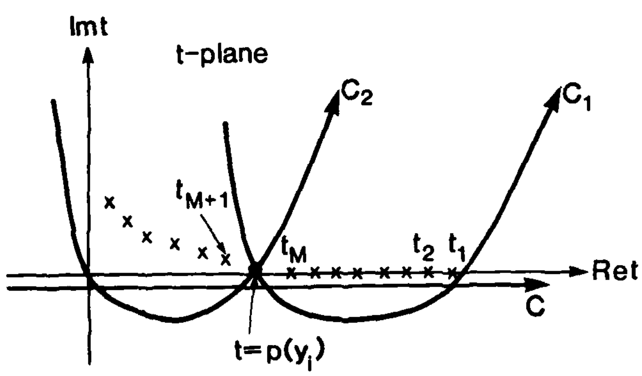


Fig. 6

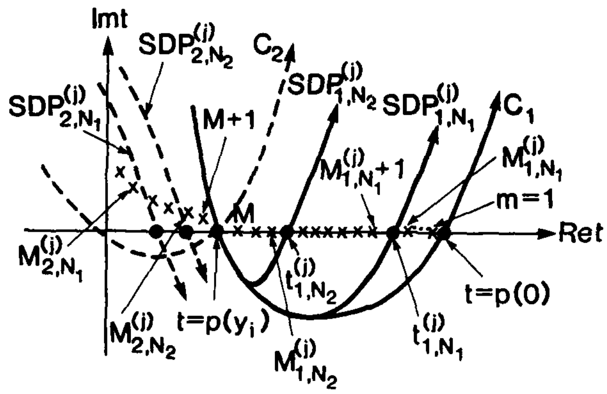


Fig. 7

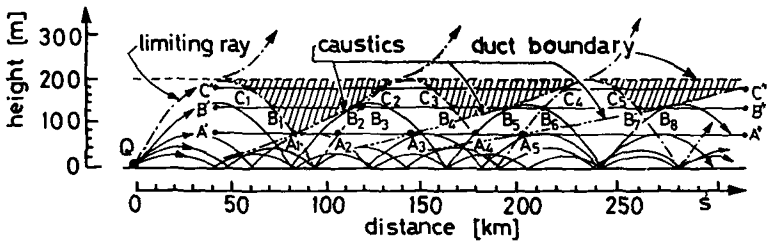


Fig. 8

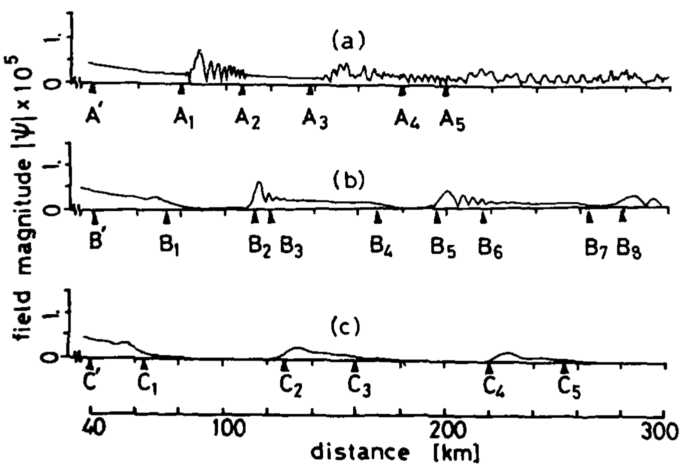


Fig. 9

The Parabolic Equation Approach to Predicting Tropospheric Propagation Effects in Operational Environments

G. Daniel Dockery
Eric R. Thews

The Johns Hopkins University Applied Physics Laboratory
Johns Hopkins Road
Laurel, Maryland 20707

ABSTRACT

The Electromagnetic Parabolic Equation (EMPE) propagation program, which was developed at The Johns Hopkins University Applied Physics Laboratory (APL), has been demonstrated to predict accurately the propagation of electromagnetic energy in environments with complicated refractivity characteristics. The validation and application of EMPE have also necessitated the development of new atmospheric measurement techniques as well as methods for representing realistically the range variations of refractivity structures. In this paper, a brief description of the parabolic equation approach and solution technique is first presented; comparisons with other propagation models are included. The measurement techniques that have been used to obtain accurate, high-resolution refractivity data are then discussed. Examples of how such refractivity data, in conjunction with a refractivity range-interpolation algorithm, have been used in EMPE to obtain good agreement with measured signal levels are also presented. The problem of developing a useful decision aid for operational systems based on the EMPE program is then addressed. The issues that are discussed include vertical and horizontal resolution of atmospheric data, timeliness of measurements, speed of propagation calculations, use of refractivity models, and automation of the environmental assessment process. An approach that is being pursued at APL for AEGIS shipboard applications, involving the use of expendable rocketborne radiosondes, is described.

1. INTRODUCTION

Open-ocean and coastal marine environments frequently contain vertical and horizontal refractive index variations that have a profound impact on the operation of radiating shipboard systems operating at frequencies above a few hundred megahertz (Refs. 1-3). In nonstandard refractive conditions, communications may be interrupted or suffer increased interference, while radar systems can experience extended or reduced target detection ranges, increased surface clutter, and altered altitude estimates. Furthermore, these effects have been observed to vary drastically with time, season, geographical location, frequency, and transmitter/receiver geometry. Clearly, the ability to predict and compensate for these effects on board an operational ship is desirable.

The idea of obtaining timely estimates of environmental impacts on radar and communications systems performance is not new. The successful development and deployment of the Integrated Refractive Effects Prediction System (IREPS) (Ref. 4) has established the usefulness of such a capability. IREPS has been installed on selected U.S. Navy ships and used in conjunction with refractivity data obtained periodically using balloonborne radiosondes to provide estimates of radar coverage and communications performance in the measured environment. In the absence of current radiosonde measurements, IREPS can also use historical refractivity data taken from a database resident in the computer program.

Recent advances in propagation modeling and atmospheric measurement suggest that some of the limitations inherent in the IREPS/balloonsonde procedure can be avoided in future environmental assessment systems. In particular, the combination of propagation models based on a parabolic wave equation, substantial increases in the capabilities of small computers, and new techniques for obtaining high-resolution refractivity information makes it feasible to obtain significant improvements in the accuracy and timeliness of system performance predictions.

There are many potential applications of radar performance predictions for both the operation of individual sensors aboard a surface ship and the coordination of multiship operations. Optimum choices of the radar parameters that are under an operator's control, such as transmitted power or detection threshold values, may be made through assessment of the impact of these parameters on total system performance in the current environment. Depending on the sophistication and flexibility of the radar system, other characteristics that may be varied include the radar waveforms used for clutter rejection and the level of sensitivity time control, which is commonly used to desensitize a radar at close ranges. Also, improvements in both low-altitude detection performance and monopulse altitude estimates may be possible through dynamic antenna beam shaping and/or pointing as a function of the environment. If optimum radar parameter choices can be established for a specified propagation environment, then integration of the decision aid process into the radar system may ultimately be desired.

Predictions of radar and weapon system coverage in the current environment can also be used to determine the most effective manner in which to position ships with respect to each other. Similarly, airborne assets can be optimized by placing surveillance aircraft in regions where detection and tracking performance with the surface-based system is predicted to be inhibited by the propagation conditions; this has been a common application of IREPS in the past. Finally, electronic support measure considerations, such as the ranges at which one's sensors may be detected, may also be addressed.

With the advent of high-fidelity propagation models and the portable computation power necessary to utilize them, the primary difficulties encountered in designing an accurate environmental assessment system are associated with acquiring the necessary atmospheric data. The most serious difficulties in this area are encountered when one decides to include horizontally varying refractivity information, because this necessitates determining the atmospheric conditions in range and azimuth as well as altitude. Potential methods of obtaining horizontally varying information include frequent flights by aircraft equipped with atmospheric sensors or dropsondes and using three-dimensional remote-sensing techniques involving radars or laser systems. Currently, these approaches are plagued with either logistic or theoretical problems. If obtaining atmospheric data in lateral as well as vertical directions proves to be impractical, one can consider using information obtained from a historical database for the current area (if one exists) or from a mesoscale forecasting model, such as those that are based on planetary boundary layer modeling (Ref. 5). Information acquired from these sources would, of course, be inferior to that obtained from *in situ*, high-resolution atmospheric measurements.

An additional goal for new environmental assessment systems is that the entire process, from the initiation of atmospheric measurements through the formation of system performance predictions, be executed rapidly and frequently enough to be applicable to the current environment. In temperate, coastal regions, diurnal variations can be severe enough to require that measurements be made as often as once an hour to achieve this goal.

The question of whether or not to attempt to account for horizontal variations and rapid temporal changes hinges on the degree of accuracy that is desired in the system performance predictions. This, in turn, depends on the applications, such as those discussed above, that are envisioned for the performance predictions. The usefulness and feasibility of such high-fidelity predictions in an operational application have yet to be investigated.

In this paper, a shipboard decision aid system that has been developed at APL for use on AEGIS ships is described. Propagation calculations for this system are performed by the EMPE program, which was developed at APL (Refs. 2 and 6). A system using a low-cost, expendable

rocketsonde, also developed at APL (Ref. 7), is used to collect atmospheric data for use in EMPE. Although EMPE can predict propagation in range-varying environments, the current version of the decision aid system assumes that the refractive conditions are horizontally homogeneous. In the sections that follow, EMPE and the two atmospheric measurement systems (including the rocketsonde) that are most commonly used by APL are described. Finally, a discussion of the decision aid system itself and examples of its output are presented.

2. PROPAGATION CALCULATIONS USING THE PARABOLIC EQUATION

Parabolic equations were used to describe electromagnetic propagation in a vertically stratified troposphere by Fock in 1946 (Ref. 8). In 1973, Tappert used the parabolic equation and an efficient numerical method, called the Fourier split-step algorithm, to model acoustic propagation in the ocean (Refs. 9 and 10). Tappert demonstrated the ability of this approach to predict propagation losses in the presence of vertically and horizontally varying sound-speed profiles. The central equation used in EMPE for electromagnetic calculations is identical to that used in the acoustic problem.

The advantage of the parabolic equation/Fourier split-step approach is that it is a full-wave, forward-scatter calculation that relies on a stable numerical scheme rather than on separation of variables or some other approach that involves simplifying approximations. As such, it is not restricted to particular regions in space (aside from avoiding the near field of the antenna), and it will account correctly for complicated variations of the refractive index in both range and altitude.

A discussion of the derivation of the scalar parabolic wave equation beginning with the vector wave equation is presented in Ref. 2 and, in less detail, in Ref. 6. Briefly, the vector wave equation describing propagation over a spherical earth can be reduced to a parabolic equation through a series of approximations that neglect the backscattered field and restrict the size of refractive index gradients that can be modeled. The refractive index gradients that violate the latter restriction are rarely encountered in the troposphere.

The problem is then translated to a pseudorectangular coordinate system, where the correction for the earth's curvature appears as a simple modification to the refractive index. The resulting parabolic equation, combined with an impedance boundary condition at the earth's surface, is an initial value problem in which the initial solution is chosen to provide the desired antenna parameters such as pattern shape, pointing direction, and antenna altitude. The solution is advanced (or "marched") in range using the Fourier split-step algorithm mentioned earlier.

In the past few years, the basic EMPE program has been upgraded to account for both ideal and measured antenna patterns and surface roughness (Refs. 6 and 11). With these added features, EMPE includes the effects of transmitter/receiver geometry, frequency, polarization, antenna pattern, the earth's curvature, electrical properties of the earth's surface, surface roughness, and variations in the refractive index.

In the examples that follow, it is convenient to describe refractive conditions using "modified refractivity" (M), usually presented as a profile versus altitude. M is related to the refractive index according to

$$M = (n - 1 + z/a) \times 10^6 , \tag{1}$$

where z is the altitude coordinate, a is the earth's radius, and n is the usual refractive index. Negative vertical gradients of M are associated with atmospheric ducts or trapping layers. Whether energy is trapped by a given "ducting" layer depends on the angle of arrival of the "rays" and on the thickness of the layer relative to the electromagnetic wavelength.

Calculations using EMPE have been compared with those generated by other models for relatively simple atmospheric conditions (Ref. 6). In Fig. 1, EMPE calculations of power relative to free space for standard "4/3 earth" atmospheric conditions (corresponding to an M-gradient of 117 per km) are compared with an effective earth radius model described by Kerr (Ref. 12); results are shown for both horizontal and vertical polarization. A frequency of 3 GHz, an antenna altitude of 31 m, and a smooth, finitely conducting sea (20° C and 3.6% salinity) are assumed, and the altitude of the calculations is 305 m. EMPE and the 4/3 earth model are observed to agree quite well, and the expected differences in interference null depths for the two polarizations are evident.

EMPE results have also been compared with those generated using a waveguide mode formalism (Refs. 13 and 14). Practically, the waveguide approach is applicable for simple refractivity profiles that are homogeneous in range and in regions that are beyond of the interference (multipath) region. A homogeneous surface duct was modeled using a refractivity gradient of -355 M/km up to an altitude of 37 m and 117 M/km above this point.

In Fig. 2, EMPE and waveguide results are plotted as vertical profiles of power relative to free space from 0 to 100 m at ranges of 40, 80, 120, and 160 km. As in the previous example, the frequency is 3 GHz and the antenna altitude is 31 m. Once again, EMPE exhibits excellent agreement with the other propagation model.

Because the parabolic equation/Fourier split-step approach can perform calculations under realistically complicated refractive conditions that are beyond the capabilities of other models, complete validation of the method is most effectively achieved via comparisons with measured data. For this reason, considerable effort has gone into collecting propagation data in situations where the refractivity environment has been characterized by meteorological measurements (Ref. 11). This has been accomplished by recording signal levels from a calibrated beacon at a coastal radar site while an instrumented helicopter collects atmospheric data along the direction in which the radar is pointed. The helicopter executes a sawtooth flight path in altitude to collect data between 3 and 300 m nominally. This procedure, which is performed over the sea, is illustrated schematically in Fig. 3. The atmospheric measurements are discussed in more detail in the next section.

One such test, performed at the NASA Wallops Flight Facility on Virginia's eastern shore, used C- and S-band shorebased radars and the corresponding transponders mounted on a twin-engined airplane that flew at a constant altitude of 31 m. The refractivity profiles measured on the afternoon of October 9, 1986, are presented in Fig. 4; the observed range dependence is not unusual for the Wallops Island coastal environment.

Range-varying conditions such as those shown in Fig. 4 usually necessitate the use of an algorithm that determines how the refractivity structures in the profiles should evolve in range. A routine that performs this function by establishing a hierarchy of refractivity structures (or layers) and matching them in adjacent profiles has been developed and is described in Ref. 11. Operating on the profiles of Fig. 4 with this algorithm and using the interpolated profiles in EMPE, the results given in Fig. 5 were obtained; the three measured data curves correspond to the three passes the airplane executed during the period when the helicopter was collecting atmospheric data. Standard atmosphere results are included for reference. The agreement obtained in this case is typical of what has been observed throughout the EMPE validation effort.

3. ATMOSPHERIC MEASUREMENTS

In designing and conducting propagation experiments for the purpose of validating EMPE's predictive capability, a number of new atmospheric measurement techniques were developed by members of APL's Space Department (Ref. 7). The two measurement systems that have been most heavily used are an instrumented Bell Jet Ranger helicopter and rocketborne radiosondes (rocketsondes). In both methods, temperature,

pressure, and humidity are recorded and subsequently converted to profiles of modified refractivity, M. The helicopter system uses a radar altimeter to obtain accurate altitude measurements, while the rocket system relies on pressure measurements to determine the altitude.

Experience gained in propagation experiments such as those described above has demonstrated that relatively high vertical resolution in atmospheric measurements is crucial. The 50- to 150-m resolution that has been obtained historically using balloon radiosondes (balloonsondes) is too large to provide an accurate representation of low-altitude refractivity structures. In most cases, for instance, surface ducts with heights of 100 m or less will go undetected.

In shipboard applications, two factors unite to reduce further the quality of balloonsonde data. First, the lowest data point (which can practically be no lower than the launch deck height) is often corrupted by the impact of the ship on the local environment. Second, in many cases, data points from the balloonsonde record are retained only if they meet some meteorological criteria; this process causes further loss of resolution and accuracy, thus making the resolution numbers quoted above optimistic.

The helicopter and rocket-based systems both provide a vertical resolution far in excess of what is typically obtained with balloonsondes. Figure 6 provides a comparison between refractivity data collected with the instrumented helicopter and that obtained with a standard balloonsonde; the data were collected simultaneously at the Wallops Island facility. Except for the surface value, which is usually very questionable for balloonsondes since there is no airflow through the sensing instrument, the agreement between the two systems is good. The profile inferred from the measurements, however, is radically different below 300 m. For a surface-based radar or communication system, of course, the conditions below 300 m are the most important.

The helicopter has been the principal platform for collecting atmospheric data during APL propagation experiments because of its ability to sample the environment in range. Using a radar altimeter, the vertical resolution of the helicopter measurements is approximately 1.0 m (Ref. 7). The range resolution that is achieved is typically on the order of 2 to 4 km; this lateral resolution has thus far proven to be adequate to characterize the environment.

The rocketsonde system operates by ejecting a radiosonde package on a parachute at an altitude that is determined by the size of the rocket engine used; the maximum altitude is generally chosen to be between 500 and 1000 m. In a shipboard application, the rocketsonde package telemeters data to a receiver until it reaches the sea surface some distance from the ship. The vertical resolution is typically 4 m or less. Comparisons between rocketsonde and helicopter measurements have consistently yielded good agreement.

A final issue must be addressed when attempting to characterize the atmosphere at very low altitudes. Despite the use of techniques that provide good vertical resolution near the ocean's surface, rapid temporal fluctuations in this region (i.e., below 20 m) often make measurements at these altitudes unstable. That is, a single profile in this very low region does not necessarily provide information on the average condition in space or time that is experienced by electromagnetic waves propagating along the sea surface.

The difficulties associated with measuring directly atmospheric quantities near the sea surface have long been recognized by investigators at the Naval Ocean Systems Center. As a result, a meteorological "evaporation duct" model, which is driven by measurements of wind speed, humidity (at an altitude of a few meters), and the air-sea temperature difference, has often been used to provide the refractivity profile at low altitudes (Ref. 15); this type of model is included in IREPS (Ref. 16). The model is intended to account for the ducting layer that occurs at the air-sea interface as a result of the rapid change in humidity that exists near the ocean's surface. The ducts are often too small to detect easily, even if stability were not a problem, but are large enough to have a significant impact on propagation, particularly for frequencies at or above 3 GHz.

An attractive option is to augment refractivity measurements made with a high-resolution technique with evaporation duct predictions using the model mentioned above. However, although the evaporation duct model has been demonstrated to explain generally the propagation statistics of long-term observations (Ref. 17), its usefulness in short-term propagation predictions has yet to be established. This issue is expected to receive considerable attention in the near future.

4. SHIPBOARD ENVIRONMENTAL ASSESSMENT USING ROCKETSONDES AND THE PARABOLIC EQUATION

The development of an operational decision aid system for use on AEGIS ships has been underway at APL for the past year. The system combines a streamlined version of the EMPE program, atmospheric data collected by rocketsondes, and a radar detection model to provide *in situ* estimates of detection performance with reasonable timeliness. In addition to using the parabolic equation to perform propagation calculations, the primary advantages of this system are that it uses high-quality refractivity data and that these data are transferred automatically from the rocketsonde receiver to the EMPE portion of the system. Persons familiar with the usual procedure associated with entering balloonsonde data into IREPS will understand the significance of the latter feature.

A version of EMPE with fixed radar and problem-size parameters has been installed on a Hewlett Packard 9020 desktop computer; this computer has superseded the HP 9845 desktop for Navy shipboard applications. It also contains the shipboard radar detection model, as well as ship stationing scenarios and engagement modeling capabilities which are beyond the scope of this article. An IBM-compatible personal computer (henceforth referred to as the IBM) does a small amount of processing on the refractivity data before their use in EMPE. The major pieces of the environmental assessment system as installed aboard an AEGIS ship are illustrated in Fig. 7.

The current version of the rocket system uses components similar to those used in hobby rockets; it is 66 cm long, 6.4 cm in diameter, and weighs 453 g. The radiosonde portion of the rocketsonde weighs 113.4 g. A new rocket design with a diameter of less than 3 cm is under development.

The radiosonde receiver/decoder (Fig. 7) demodulates the narrowband FM signal received from the rocketsonde instruments and performs frame synchronization and error correction on the data. As the data are received, the IBM, running proprietary software, scales the data to account for instrument calibrations and stores the raw data on disk. After the parachuting package has impacted the water, a data conditioning program in the IBM scans the raw data for the maximum altitude achieved and discards bad data (which may result from telemetry dropouts, unrecoverable errors in the telemetry code, or erroneous measurements). Using the remaining data, a modified refractivity versus altitude file is created, plotted on the computer screen, and subsequently formatted and transmitted over a fiberoptic link to the HP 9020 computer located in the Combat Information Center of the AEGIS ship. Upon receipt of the data, EMPE automatically begins its calculations, and users of the decision aid system are informed that new propagation data will soon be available.

The preparation of the rocket and instruments before launch takes about 10 min, and an average of 6 min is used in the period beginning with the launch and ending with the splash of the instrument package. Raw data conditioning and transfer over the fiberoptic link require approximately 4 min, and EMPE uses 17 min to complete a calculation out to approximately 120 km in 0.46-km range steps. Thus, new propagation data are available for further processing by the decision aid system 27 min after a rocket launch.

To limit the execution of the propagation portion of the program to a reasonable length of time, the altitudes of the EMPE calculation are restricted to 0 to 305 m; precalculated 4/3-earth propagation loss values are used for higher altitude performance predictions. The examples presented below are restricted to the region for which EMPE calculations have been performed.

The low-altitude radar model incorporates the appropriate parameters such as power, transmit and receive antenna gains and losses, processing gains and losses, and radar operator definable options. Sensitivity time control and barrage jamming effects are also included. The procedure involves using the EMPE propagation loss values in calculations of two-way power at positions on the EMPE mesh; the full resolution in range is retained, and the vertical resolution is nominally 1.5 m. The two-way power and other radar parameters are then used to determine the target radar cross section required for a 50% probability of detection at each point on the mesh. By assigning colors or shades of gray to specified ranges of target cross section, contours of 50% probability of detection are generated.

The decision aid system described above was installed and tested at sea on an AEGIS ship in November 1988. A profile of modified refractivity derived from data collected by a rocketsonde during this test is presented in Fig. 8; these data were collected approximately 180 km off the east coast of the United States. The slope of the profile is 121.4 *M*/km, which is very close to the 4/3-earth gradient.

The propagation and detection calculations presented here were performed for a hypothetical C-band radar with an antenna altitude of 31 m, a beamwidth of 2°, and a pointing direction of 1°. The transmitted power and other system parameters correspond to a free space detection range of 150 km for a 1-m^2 target.

Figure 9 presents a gray-scale contour of 50% probability of detection for the hypothetical radar system using the profile of Fig. 8; this is a typical detection contour for standard atmosphere propagation conditions. Refractivity data collected the day before those shown in Fig. 8 are given in Fig. 10; a small ducting layer is evident below 35 m. The corresponding detection contour is presented in Fig. 11. The impact of the duct is clear; greatly increased detection ranges are indicated for targets in the duct. Above the duct, detection probabilities are enhanced at all altitudes for larger targets and reduced somewhat for smaller targets.

5. CONCLUSION

In the November 1988 at-sea tests, the EMPE/rocketsonde assessment system demonstrated a capability to provide low-altitude radar performance predictions in a timely and virtually automated manner. The parabolic equation method of modeling propagation is the only currently available technique for accurately predicting relative power levels both within and beyond the horizon in complicated refractivity conditions. Furthermore, the use of rocketsondes, which provide high-resolution atmospheric data all the way down to the ocean's surface, represents a substantial improvement in data quality when compared with balloonsonde systems.

As computer technology continues to advance, it will become increasingly practical to extend shipboard parabolic equation calculations to higher altitudes and longer ranges; the HP 9020 machine used thus far is, in fact, considerably slower than state-of-the-art computers of that size. The use of a small, but dedicated, array processor would also significantly decrease the required execution time.

The largest potential difficulty is associated with the acquisition of laterally varying atmospheric data. In the absence of a proven remote-sensing technique involving high-resolution (i.e., laser) radars or similar approaches, sounding the atmosphere in three dimensions presents enormous logistic problems. On the other hand, it seems unlikely that a high level of confidence can be associated with performance predictions generated using a single refractivity profile, particularly in temperate coastal regions. Therefore, future efforts should include studies to determine the value of representing horizontally varying propagation conditions in shipboard applications.

6. ACKNOWLEDGMENTS

Many people at APL have been involved in the development of propagation prediction and environmental assessment capabilities. The credit for initial application of the parabolic equation/Fourier split-step approach to electromagnetic tropospheric propagation belongs to H. Ko, J. W. Sari, J. P. Skura, and R. I. Joseph. G. C. Konstanzer is responsible for developing the refractivity interpolating algorithm that has made it possible to perform EMPE calculations using complicated, range-varying atmospheric data.

The helicopter and rocketsonde measurement systems were developed and operated by J. R. Rowland and S. M. Babin under the supervision of J. Goldhirsh. Mr. Rowland and Mr. Babin were also instrumental in adapting the rocketsonde equipment for use in the decision aid system. Finally, R. G. Roll and G. E. Frishkorn were responsible for installing EMPE on the HP 9020 and integrating EMPE and the radar model with the existing decision aid program, respectively. This work was supported by the Naval Sea Systems Command, PMS-400B.

7. REFERENCES

1. H. V. Hitney, J. H. Richter, R. A. Pappert, K. D. Anderson, and G. H. Baumgartner, Jr., "Tropospheric Radio Propagation," *Proc. IEEE*, Vol. 73, No. 2, Feb. 1985, pp. 265–283.

2. H. W. Ko, J. W. Sari, and J. P. Skura, "Anomalous Microwave Propagation Through Atmospheric Ducts," *Johns Hopkins APL Tech. Dig.*, Vol. 4, No. 1, 1983, pp. 12–26.

3. R. A. Pappert and C. L. Goodhart, "Case Studies of Beyond-the-Horizon Propagation in Tropospheric Ducting Environments," *Rad. Sci.*, Vol. 12, No. 1, 1977, pp. 75–87.

4. H. V. Hitney and J. H. Richter, "Integrated Refractive Effects Prediction System (IREPS)," *Nav. Eng. J.*, Vol. 88, 1976, pp. 257–262.

5. C. E. Schemm, L. P. Manzi, and H. W. Ko, "A Predictive System for Estimating the Effects of Range- and Time-Dependent Anomalous Refraction on Electromagnetic Wave Propagation," *Johns Hopkins APL Tech. Dig.*, Vol. 8, No. 4, 1987, pp. 394–403.

6. G. D. Dockery, "Modeling Electromagnetic Wave Propagation in the Troposphere Using the Parabolic Equation," *IEEE Trans. Ant. Propag.*, Vol. 36, No. 10, 1988, pp. 1464–1470.

7. J. R. Rowland and S. M. Babin, "Fine-Scale Measurements of Microwave Refractivity Profiles with Helicopter and Low-Cost Rocket Probes," *Johns Hopkins APL Tech. Dig.*, Vol. 8, No. 4, 1987, pp. 413–417.

8. V. A. Fock, "Solution of the Problem of Propagation of Electromagnetic Waves Along the Earth's Surface by Method of Parabolic Equations," *J. Phys. USSR*, Vol. 10, No. 1, 1946, pp. 13–35.

9. F. D. Tappert, "The Parabolic Approximation Method," in *Wave Propagation and Underwater Acoustics*, J. B. Keller and J. S. Papadakis, eds., New York, Springer-Verlag, 1977.

10. F. R. DiNapoli and R. L. Deavenport, "Numerical Methods of Underwater Acoustic Propagation," in *Ocean Acoustics*, J. A. DeSanto, ed., New York, Springer-Verlag, 1977.

11. G. D. Dockery and G. C. Konstanzer, "Recent Advances in Prediction of Tropospheric Propagation Using the Parabolic Equation," *Johns Hopkins APL Tech. Dig.*, Vol. 8, No. 4, 1987, pp. 404–412.

12. D. E. Kerr, *Propagation of Short Radio Waves*, New York, McGraw-Hill, 1951, Ch. 2 and 5.

13. S. W. Marcus, "A Model to Calculate EM Fields in Tropospheric Duct Environments at Frequencies Through SHF," *Rad. Sci.*, Vol. 17, No. 5, 1982, pp. 895–901.

14. G. B. Baumgartner, Jr., "XWVG: A Waveguide Program for Trilinear Tropospheric Ducts," 1983, Technical Document 610, Naval Ocean Systems Center, San Diego, Cal.

15. H. Jeske, "Die Ausbreitung elektromagnetischer Wellen im cm-bis m-Band uber dem Meer unter besonderer Berucksichtigung der meteorologischen Bedingungen in der maritimen Grenzshicht," in *Hamburger Geophysikalische Einzelschriften*, De Gruyter, Hamburg, 1965.

16. R. A. Paulus, "Practical Application of an Evaporation Duct Model," *Rad. Sci.*, Vol. 20, No. 4, 1985, pp. 887-896.

17. C. P. Hattan, "Evaporation-Duct Effects on the Detection of Low-Altitude Targets," Jan 1988, Technical Report 1201, Naval Ocean Systems Center, San Diego, Cal.

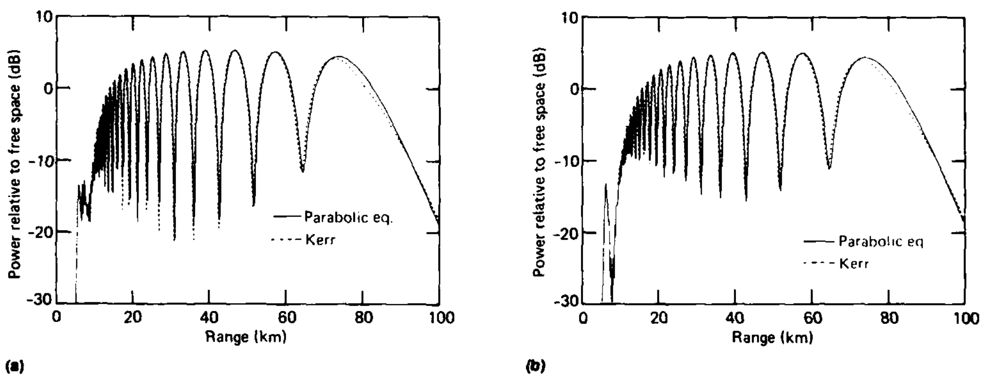


Figure 1. Results from EMPE and a standard atmosphere model for a frequency of 3 GHz, an altitude of 305 m, and an antenna altitude of 31 m; (a) horizontal polarization and (b) vertical polarization.

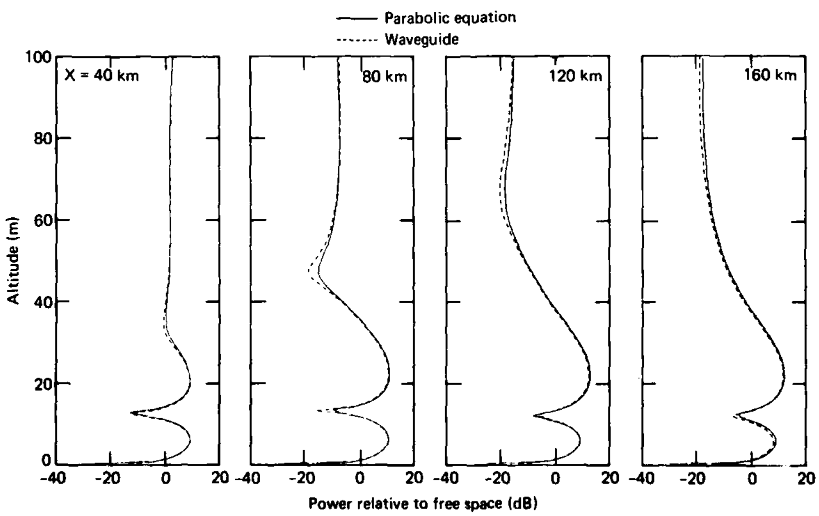


Figure 2. Comparison of EMPE and waveguide mode results for 3-GHz horizontal polarization in a uniform surface duct.

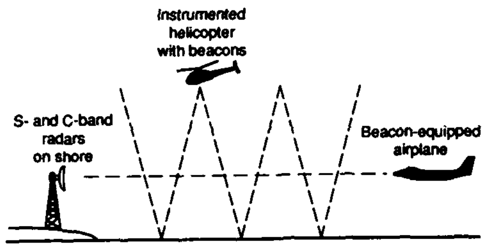


Figure 3. Typical arrangement for propagation experiments conducted at Wallops Island, Virginia.

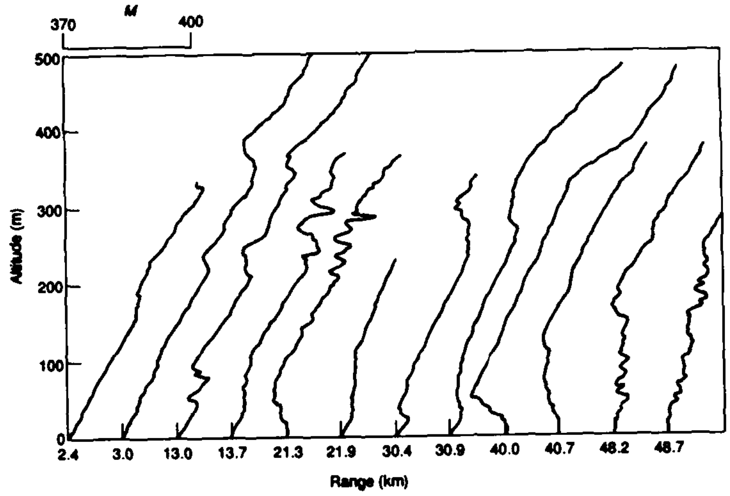


Figure 4. Refractivity profiles derived from helicopter measurements taken at Wallops Island on October 9, 1986.

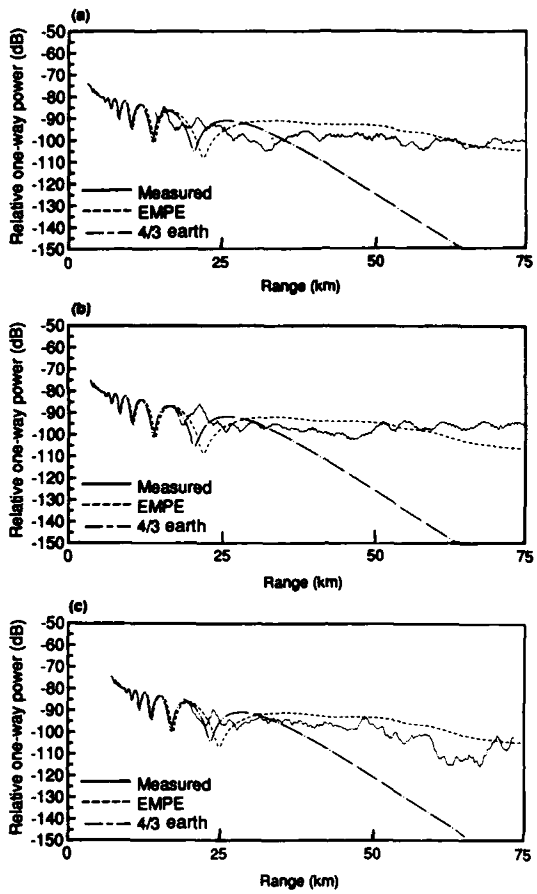


Figure 5. EMPE-generated and observed C-band signal levels measured during three level flights by a beacon-equipped airplane, October 9, 1986.

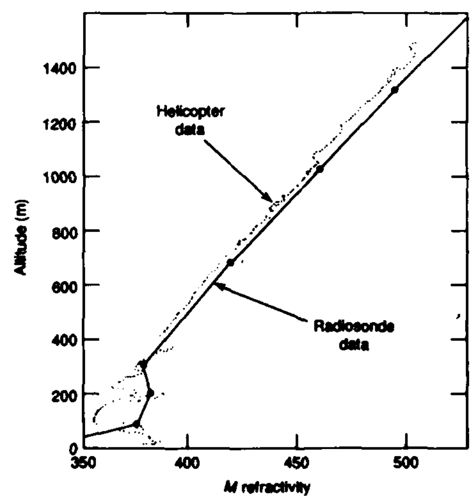


Figure 6. Comparison of simultaneous balloonsonde and helicopter refractivity measurements conducted at Wallops Island.

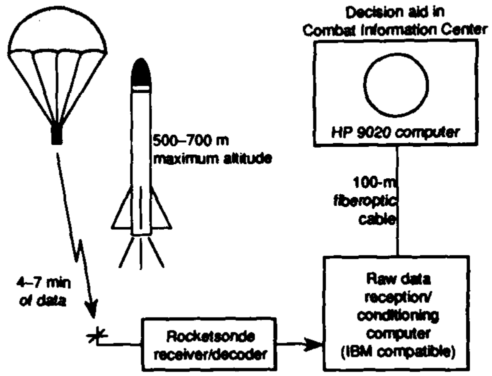


Figure 7. Schematic diagram of environmental assessment system.

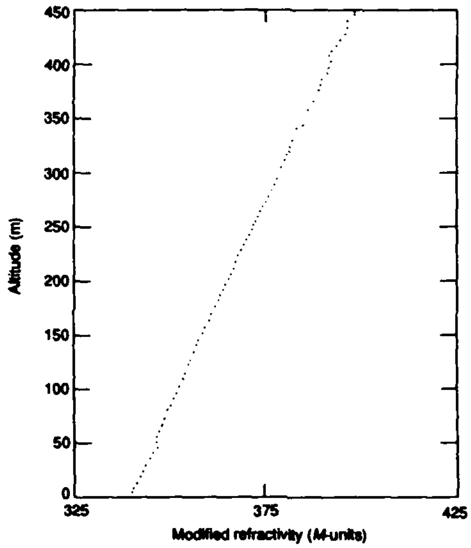


Figure 8. Standard atmosphere rocketsonde profile obtained from an AEGIS ship in November 1988.

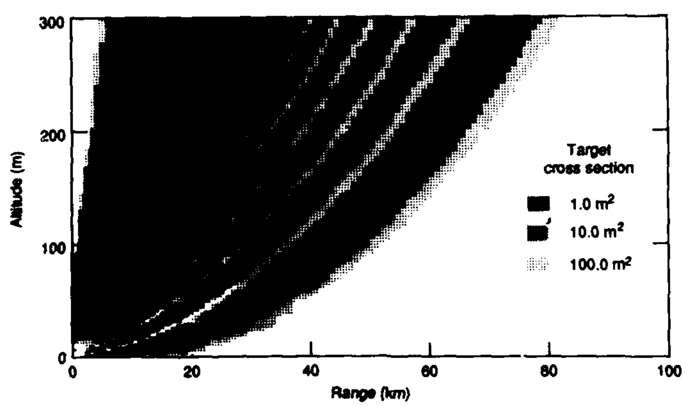


Figure 9. Gray-scale contour of 50% probability of detection for a hypothetical C-band radar system and the rocketsonde profile of Figure 8.

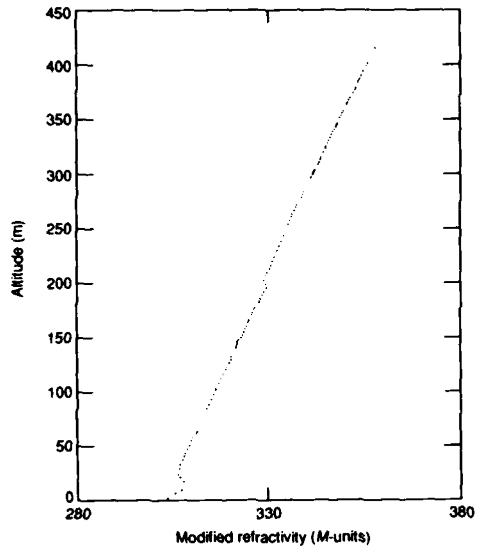


Figure 10. Ducting rocketsonde profile obtained from an AEGIS ship in November 1988.

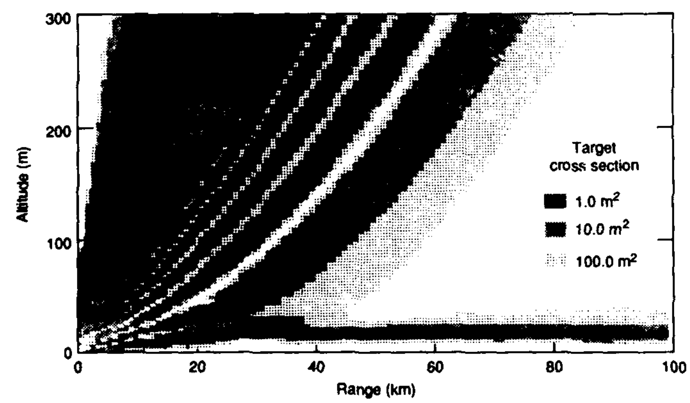


Figure 11. Gray-scale contour of 50% probability of detection for a hypothetical C-band radar system and the rocketsonde profile of Figure 10.

RPE: A Parabolic Equation Radio Assessment Model

F.J. RYAN

Ocean and Atmospheric Sciences Division
CODE 541
Naval Ocean Systems Center
San Diego, CA 92152-5000
USA

SUMMARY

The use of parabolic wave equation codes to model tropospheric radio propagation is rapidly gaining popularity. To properly implement these powerful techniques into radar performance assessment models, requires a detailed understanding of the inherent errors in the parabolic approximation. This paper examines three major sources of error in split-step parabolic equation solvers: 1) approximation of the elliptic wave operator by a parabolic operator, 2) truncation error in the formal parabolic solution, and 3) truncation error in the split-step operator. These errors are discussed in the context of operational codes.

1. INTRODUCTION

In December 1901, Marconi became the first to transmit radio signals across the Atlantic Ocean thereby sparking interest in the propagation of electromagnetic waves over the surface of the earth. In the following forty years, much progress was made by mathematicians and physicists in understanding low frequency radio propagation. With the development of radars during WWII, it became readily apparent that refraction and diffraction phenomena at microwave frequencies were driven by environmental variations in the troposphere. While advances were made in understanding propagation in vertically stratified atmospheres using normal mode methods, the problem of horizontal environmental variations remained unsolved. In 1946 Leontovich and Fock[1,2] made a breakthrough when they introduced the concept of a parabolic wave equation (PE) to model transhorizon radio propagation. However, it was almost thirty years before a practical solution to the Leontovich-Fock parabolic wave equation was developed by Hardin and Tappert[3] to model ionospheric propagation. The Tappert-Hardin split-step Fourier method exploited advances in computer hardware and the development of fast Fourier transform (FFT) algorithms to yield efficient numerical solutions to the Leontovich-Fock parabolic wave equation. In 1974 Tappert[4] introduced split-step PE to the underwater acoustics community and it rapidly became a valuable tool for predicting range dependent underwater sound propagation. In 1983 the split-step PE method was re-applied to radar propagation by Ko *et al*[5] to study anomalous tropospheric propagation. Since then a number of other groups have developed radar PE models including the Naval Ocean Systems Center (NOSC).

The NOSC radio parabolic equation (RPE) code is based upon the split-step Fourier algorithm solution to the parabolic wave equation and incorporates many features designed to make it useful for radar system performance modeling. The goal was to design a robust PE code which is highly automated, that incorporates local step-wise error control, and is very efficient in terms of computer resources. A good deal of effort was directed toward ensuring the fidelity of model outputs—an ever present problem when dealing with numerical codes.

Because the application of split-step PE methods to tropospheric radar problems has become commonplace, the formal development of the algorithm will only be briefly described. Rather, this paper will focus on some of the more subtle issues which arise in implementing a numerical PE code; namely the control of truncation errors and techniques for dealing with complicated boundary conditions. Too often, the degree to which error controls are implemented in numerical codes, such as PE, can make the difference between a viable computer model and one in which the outputs are in question. In section 2, an exact "earth-flattened" elliptic cartesian wave equation is derived which describes electromagnetic wave propagation in a spatially inhomogeneous troposphere. Operator splitting methods are then used to derive a parabolic wave equation. Section 3 deals with an error analysis of the split-step method.

2. PARABOLIC APPROXIMATION

In this section the elliptic Helmholtz equation which governs the propagation of linearly polarized radiation over the earth's surface is derived from Maxwell's equations. The resulting two-dimensional elliptic partial differential equation in earth-centered coordinates is then mapped exactly into a cartesian frame and approximated by a parabolic system.

Elliptic Wave Equation

The problem at hand is to describe mathematically electromagnetic wave propagation in a spatially inhomogeneous troposphere. Rationalized mks units are employed and monochromatic radiation having an implicit time dependence $\exp(-i\omega t)$, where ω is the radian frequency, is assumed. A linearly polarized source is assumed located at $\vec{R}_0 = (r_0, 0, 0)$ in an earth-based spherical coordinate system (r, θ, ϕ) with polar axis co-linear with the source. The propagation medium is assumed to vary only in the meridonal

plane connecting the source and receiver (i.e. no ϕ dependence). Under these assumptions and using standard methods to convert the first order Maxwell equations to a second order system, the spectral component of the magnetic intensity vector, $\mathbf{H}(\vec{r},\omega)$, can be shown to satisfy

$$\nabla^2\mathbf{H}(\vec{r}) - \nabla\times\mathbf{H}(\vec{r})\times\frac{\nabla\epsilon(\vec{r})}{\epsilon(\vec{r})} + \omega^2\mu_0\epsilon(\vec{r})\mathbf{H}(\vec{r}) = 0, \tag{1}$$

where μ_0 is the free-space magnetic permeability, $\epsilon(\vec{r})$ is the absolute dielectric constant, and $\vec{r} = r\hat{e}_r + \theta\hat{e}_\theta$ is the position vector in the source-receiver plane. Similarly, the electric field vector $\mathbf{E}$ can be shown to satisfy an equation of the form

$$\nabla^2\mathbf{E}(\vec{r}) + \nabla\left(\mathbf{E}(\vec{r})\cdot\frac{\nabla\epsilon(\vec{r})}{\epsilon(\vec{r})}\right) + \omega^2\mu_0\epsilon(\vec{r})\mathbf{E}(\vec{r}) = 0. \tag{2}$$

Since the equation satisfied by the electric field is contained in Eq. (1), attention will be focused on obtaining solutions to the latter.

If the source emits vertically polarized radiation, then the non-zero magnetic field vector components are $\mathbf{H} = H_\phi(\vec{r})\hat{e}_\phi$ and Eq. (1) becomes in spherical coordinates

$$\frac{\epsilon(\vec{r})}{r}\frac{\partial}{\partial r}\frac{1}{\epsilon(\vec{r})}\frac{\partial rH_\phi(\vec{r})}{\partial r} + \frac{\epsilon(\vec{r})}{r^2\sin\theta}\frac{\partial}{\partial\theta}\frac{\sin\theta}{\epsilon(\vec{r})}\frac{\partial H_\phi(\vec{r})}{\partial\theta} + \left[\omega^2\mu_0\epsilon(\vec{r}) - \frac{\cot\theta}{r^2\epsilon(\vec{r})}\frac{\partial\epsilon(\vec{r})}{\partial\theta}\right]H_\phi(\vec{r}) = 0. \tag{3}$$

Equation (3) is not in the desired Helmholtz form but can be transformed into it by a slight change of variable. Define a new dependent field variable u by

$$u(\vec{r}) \equiv u(r,\theta) = \frac{\sqrt{\sin\theta}}{n(r,\theta)}rH_\phi(r,\theta), \qquad n(r,\theta) = \sqrt{\epsilon(r,\theta)/\epsilon_0}, \tag{4}$$

where n is the index of refraction referenced to the vacuum dielectric constant ϵ_0 . In terms of u, Eq. (3) becomes the desired Helmholtz equation:

$$\left[\frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2}{\partial\theta^2} + k^2n_{eff}^2(r,\theta)\right]u(r,\theta) = 0, \tag{5}$$

where $k = \omega\sqrt{\mu_0\epsilon_0}$ is the vacuum wavenumber, and n_{eff} is an "effective" index of refraction defined by

$$\begin{aligned} n_{eff}^2 &= n^2(r,\theta) + \frac{\delta n(r,\theta)}{k^2}, \\ \delta n(r,\theta) &= -\frac{\cot\theta}{r^2 n}\frac{\partial n}{\partial\theta} + \frac{1}{4r^2} + \frac{1}{4r^2\sin^2\theta} - \frac{n}{r^2}\frac{\partial^2 n^{-1}}{\partial\theta^2} - n\frac{\partial^2 n^{-1}}{\partial r^2}. \end{aligned} \tag{6}$$

The additional term δn in Eq. (6) is generally small except in two cases. The first case occurs for propagation near the boundary separating dielectrics with finite conductivity, in which case the spatial derivatives may get large. The second case occurs at low frequency where the wavenumber k is small. Inclusion of the δn correction allows one to deal in a unified manner with propagation over idealized boundaries as well as treat propagation in the presence of sub-surface overburden.

Exact Earth-flattening Transform

The coupling of r and θ variables in Eq. (5) is not desirable for numerical work and can be eliminated by means of a simple transformation from the spherical (r,θ) coordinate system to an equivalent cartesian (z,x) coordinate system defined by

$$\begin{aligned} z &= a\ln(1+h/a) \quad \approx h, \qquad \text{if} \quad h/a \ll 1, \\ x &= a\theta, \end{aligned} \tag{7}$$

where a is the earth's radius, and $h \equiv r - a$ is the local altitude. This transformation is commonly denoted as an "earth-flattening" transformation, and with the approximation $z \approx h$ has been studied by Pekeris[6]. In terms of the new flat-earth variables, define a modified index of refraction, m, by

$$m(x,z) = n(\theta,r)e^{z/a}, \quad \approx n(x,z)(1+z/a), \quad \text{if} \quad z/a \ll 1, \tag{8}$$

and a new dependent field variable,w, by

$$\begin{aligned} w(x,z) \equiv e^{z/a}u(x,z) &= ae^{3z/2a}\frac{\sqrt{\sin(x/a)}}{m(x,z)}H_\phi(x,z), \\ &\approx \frac{\sqrt{ax}}{m(x,z)}H_\phi(x,z), \qquad \text{if} \quad x/a \ll 1. \end{aligned} \tag{9}$$

Then, in the flat-earth coordinate system, Eq. (5) takes the form

$$\left[\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial z^2}+K^2(x,z)\right]w(x,z)=0, \tag{10}$$

where $K(x,z)$ is the "effective" wavenumber defined by

$$K^2(x,z)=k^2m^2(x,z)-\frac{\cot(x/a)}{am}\frac{\partial m}{\partial x}+\frac{\csc^2(x/a)}{4a^2}+\frac{1}{m}\frac{\partial^2 m}{\partial x^2}+m\frac{\partial^2 m^{-1}}{\partial z^2}-\frac{m}{a}\frac{\partial m^{-1}}{\partial z}. \tag{11}$$

For no range dependence of m, Eq. (11) becomes

$$\begin{aligned} K^2(x,z) &= k^2 m_{eff}^2(z)+\frac{1}{4x^2}, \qquad \text{if} \quad x/a \ll 1, \\ m_{eff}^2(z) &= n^2(z)e^{2z/a}+\frac{1}{k^2}\left[\frac{1}{n}\frac{\partial^2 n}{\partial z^2}-2\left(\frac{1}{n}\frac{\partial n}{\partial z}\right)^2-\frac{1}{an}\frac{\partial n}{\partial z}\right], \\ &\approx n^2(z)(1+2z/a), \qquad \text{if} \quad z/a \ll 1. \end{aligned} \tag{12}$$

which is the same form as the earth-flattening approximation discussed by Pekeris. The Pekeris transform, however, was based upon $z=r-a$ and is only valid for small z/a in contrast to that given by Eq. (7) which is exact.

While Eq. (10) follows exactly from Maxwell's equations, analytic solutions are only possible if K is independent of x. Nor is direct numerical solution feasible; because Eq. (10) is an *elliptic partial differential* equation, it must be numerically solved over the entire propagation domain simultaneously.

To deal with this problem, Leontovich and Fock introduced the concept of a parabolic wave equation. The approach of Leontovich and Fock was to make the envelope transformation

$$w(\vec{x})=e^{ik_0x}\psi(\vec{x}), \tag{13}$$

then substitute it into Eq. (10), and drop a term, $\frac{\partial^2\psi(\vec{x})}{\partial x^2}$, to yield the Leontovich-Fock parabolic wave equation (PE):

$$2ik_0\frac{\partial\psi(\vec{x})}{\partial x}+\frac{\partial^2\psi(\vec{x})}{\partial z^2}+\left(K^2(\vec{x})-k_0^2\right)\psi(\vec{x})=0, \tag{14}$$

where k_0 is a reference wavenumber. The justification for dropping the second derivative term is usually made on the grounds that the envelope function ψ is a slowly varying function of the range coordinate x. While pedagogically correct, such a derivation of PE does not yield any insight into the approximation nor the errors that are incurred in using it. An alternative approach to deriving the PE is via factorization of the elliptic wave equation. If we define an operator $Q(\vec{x})$ by

$$Q(\vec{x})=\sqrt{\frac{\partial^2}{\partial z^2}+K^2(\vec{x})}, \tag{15}$$

then Eq. (10) can be expressed equivalently in the factored form

$$\left(\frac{\partial}{\partial x}+iQ\right)\left(\frac{\partial}{\partial x}-iQ\right)w(\vec{x})+i\left[\frac{\partial}{\partial x},Q\right]w(\vec{x})=0. \tag{16}$$

(The notation $\left[\frac{\partial}{\partial x},Q\right]=\frac{\partial Q}{\partial x}-Q\frac{\partial}{\partial x}$ is the commutator of the operators $\frac{\partial}{\partial x}$ and Q.)

For range independent propagation $\left[\frac{\partial}{\partial x},Q\right]=0$, and the equation satisfied by outwardly propagating waves is just

$$\frac{\partial w(\vec{x})}{\partial x}=iQ(\vec{x})w(\vec{x}). \tag{17}$$

Equation (17) represents an exact formal solution to the range independent Helmholtz equation, and is the most general PE which is exact for range independent media. Following Tappert[7], a parabolic wave equation with the Q operator defined by Eq. (15) will be denoted as the general PE (GPE) and the Q operator will be denoted as the GPE propagator. GPE is the most complete PE which is evolutionary in range and neglects backscattering. For range independent environments it is exact within the limits of the far-field approximation, and is the starting point for all numerical PE algorithms. In range dependent media, however, the commutator $\left[\frac{\partial}{\partial x},Q\right]\neq 0$ and must be accounted for in solving Eq. (16).

Bremmer Series

One method for dealing with range dependence is to develop forward scatter approximations that are multi-dimensional analogs of Bremmer's[8] series solutions to second order ordinary differential equations. This method was utilized by Corones[9] to analyze elliptic wave equations. Corones split the full-wave elliptic solution w into two components: one propagating away from the source toward large x, called the transmitted of forward scattered field w^+, and the other propagating inward toward small x, the reflected or backscattered field w^-. This splitting of the total wave field into transmitted and reflected waves is effected by operators that locally take transmitted waves into reflected waves and vice versa. In terms of differential equations, the splitting operators determine the coupling between the forward scattered and back scattered wave fields—precisely what is missing in the conventional derivation of the parabolic wave equation.

Any second order system such as Eq. (16) may be recast into two coupled first order equations as

$$\frac{\partial}{\partial x}\begin{pmatrix} w(\vec{x}) \\ \dfrac{\partial w(\vec{x})}{\partial x} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -Q^2(\vec{x}) & 0 \end{pmatrix} \cdot \begin{pmatrix} w(\vec{x}) \\ \dfrac{\partial w(\vec{x})}{\partial x} \end{pmatrix}. \tag{18}$$

The splitting of $w(\vec{x})$ into transmitted and reflected wave components is defined by a splitting matrix $S(\vec{x})$, with the transmitted, w^+, and reflected, w^-, components given by

$$\begin{pmatrix} w^+(\vec{x}) \\ w^-(\vec{x}) \end{pmatrix} = S(\vec{x}) \begin{pmatrix} w(\vec{x}) \\ \dfrac{\partial w(\vec{x})}{\partial x} \end{pmatrix}. \tag{19}$$

The substitution of Eq. (19) into Eq. (18) yields coupled equations for the transmitted and reflected wave fields:

$$\frac{\partial}{\partial x}\begin{pmatrix} w^+(\vec{x}) \\ w^-(\vec{x}) \end{pmatrix} = \left[\frac{\partial S}{\partial x} S^{-1}(\vec{x}) + S(\vec{x}) \begin{pmatrix} 0 & 1 \\ -Q^2(\vec{x}) & 0 \end{pmatrix} S^{-1}(\vec{x})\right] \begin{pmatrix} w^+(\vec{x}) \\ w^-(\vec{x}) \end{pmatrix}. \tag{20}$$

Various forms of the splitting matrix S are possible; in fact, $S(\vec{x})$ is formally arbitrary. One such splitting is afforded by

$$S(\vec{x}) = \begin{pmatrix} 1 & -i/Q(\vec{x}) \\ 1 & +i/Q(\vec{x}) \end{pmatrix}, \tag{21}$$

and yields the following coupled equations for w^+ and w^-:

$$-i\frac{\partial w^+(\vec{x})}{\partial x} = \left(1 - \frac{i}{2}\frac{\partial Q^{-1}(\vec{x})}{\partial x}\right) Q(\vec{x}) w^+(\vec{x}) + \frac{i}{2}\frac{\partial Q^{-1}(\vec{x})}{\partial x} Q(\vec{x}) w^-(\vec{x}), \tag{22a}$$

$$-i\frac{\partial w^-(\vec{x})}{\partial x} = -\left(1 + \frac{i}{2}\frac{\partial Q^{-1}(\vec{x})}{\partial x}\right) Q(\vec{x}) w^-(\vec{x}) + \frac{i}{2}\frac{\partial Q^{-1}(\vec{x})}{\partial x} Q(\vec{x}) w^+(\vec{x}). \tag{22b}$$

For this definition of the splitting matrix, the above equations decouple exactly if $\frac{\partial Q}{\partial x} = 0$, and yield exact expressions for the forward and back-scattered wave fields. When $K(\vec{x})$ is a slowly varying function of range, then the parabolic approximation is obtained for the transmitted field w^+.

3. NUMERICAL SOLUTIONS

This section deals with the numerical solution of the GPE, Eq. (17), in cases where the range dependent commutator $[\frac{\partial}{\partial x}, Q] \ll 1$. Because of the square root form of the GPE propagator, normal solution techniques for parabolic systems are not possible. This is because Q belongs to the class of pseudo-differential operators since it contains both multiplicative, K^2, and differential, $\frac{\partial^2}{\partial z^2}$, operators under the radical. Hence, $Q(\vec{x})w(\vec{x})$ cannot be expressed as a finite Taylor series about the point $\vec{x}$. To resolve this problem and effect a numerical solution to PE, the Q pseudo-differential operator must be approximated by ordinary operators:

$$Q(x,z) \approx A(z) + B(x,z). \tag{23}$$

Then, a formal solution of Eq. (17) is expressed as

$$\begin{aligned} \psi(x_0 + \Delta, z) &= e^{\int_{x_0}^{x_0+\Delta}(A+B(x'))\,dx'}\,\psi(x_0, z), \\ &\approx e^{i\Delta(A+B)}\psi(x_0, z). \end{aligned} \tag{24}$$

Finally, the exponential operator in Eq. (24) is symmetrically factored into the product of exponential operators

$$\psi(x_0 + \Delta, z) = e^{i\Delta A/2}\, e^{i\Delta B}\, e^{i\Delta A/2}\psi(x_0, z), \tag{25}$$

to yield the split-step PE solution.

Split-step PE is thus predicated on three approximations: first, approximating the square root GPE propagator by a finite sum of ordinary operators, Eq. (23); second, assuming an exponential PE solution form, Eq. (24); finally, approximating the exponential PE solution operator by a symmetrized factorization of individual operators, Eq. (25). Each of these steps introduces an intrinsic error in the split-step PE solution—the ability to quantify these errors is crucial to the successful implementation of a numerical PE code.

Q Splitting

The Q operator can be expressed in terms of a "kinetic energy" operator, T, and a "potential energy" operator, V, as

$$Q(\vec{x}) = k_0\sqrt{1 + T(z) - V(x,z)}, \tag{26}$$

where $T(z) \equiv \dfrac{1}{k_0^2}\dfrac{\partial^2}{\partial z^2}$, and $V(x,z) \equiv 1 - K^2(x,z)/k_0^2$. If $T - V$ is small, then a Taylor series expansion of Eq. (26) gives

$$Q_1(\vec{x}) = k_0 + \frac{k_0}{2}(T - V), \tag{27}$$

which is the standard PE (SPE) approximation proposed by Hardin and Tappert. The Q_1 approximation to Q basically assumes that the propagation is within a small cone of angles centered about k_0. An alternative approximation to Q was developed by Feit and Fleck[10]:

$$Q_2(\vec{x}) = k_0\sqrt{1+T} + K(\vec{x}) - k_0. \tag{28}$$

The approximation Q_2 is known as a "wide-angle" parabolic approximation (WAPE) propagator since it is valid for much wider propagation angles.

It proves convenient to make the Leontovich-Fock envelope transform Eq. (13), in which case Eq. (17) becomes

$$\begin{aligned}\frac{\partial\psi(\vec{x})}{\partial x} &= i\,(Q(\vec{x}) - k_0)\,\psi(\vec{x}),\\ &\approx i(A(z) + B(x,z))\psi(\vec{x}).\end{aligned} \tag{29}$$

If the SPE propagator, Eq. (27), is used to approximate $Q(x)$ then

$$\frac{\partial\psi(\vec{x})}{\partial x} = i\left(\frac{1}{2k_0}\frac{\partial^2}{\partial z^2} - \frac{k_0}{2}V(\vec{x})\right)\psi(\vec{x}), \tag{30}$$

with

$$A(z) \equiv \frac{1}{2k_0}\frac{\partial^2}{\partial z^2}, \qquad B(x,z) \equiv -\frac{k_0}{2}V(x,z); \tag{31}$$

while if the WAPE propagator, Eq. (28), is used

$$\frac{\partial\psi(\vec{x})}{\partial x} = i\left(\frac{\frac{\partial^2}{\partial z^2}}{k_0 + \sqrt{k_0^2 + \frac{\partial^2}{\partial z^2}}} + K(\vec{x}) - k_0\right)\psi(\vec{x}), \tag{32}$$

with

$$A(z) \equiv \frac{\partial^2}{\partial z^2}\Big/\left(k_0 + \sqrt{k_0^2 + \frac{\partial^2}{\partial z^2}}\right), \quad B(x,z) \equiv K(x,z) - k_0. \tag{33}$$

The PE solution, $\psi(x_0 + \Delta, z)$, at range $x_0 + \Delta$ is obtained from the known field $\psi(x_0, z)$ by means of the split-step operator, Eq. (25). The presence of a differential operator, $A(z)$, in the split-step exponent can be dealt with by transforming to a basis in which $A(z)$ is diagonal. This can be accomplished by Fourier transforming the PE field, in which case the SPE operator $\exp(i\Delta A/2)\psi(x,z)$ is effected via

$$e^{i\Delta A/2}\psi(x,z) = \mathcal{F}^{-1}\left[e^{-i\Delta p^2/4k_0}\mathcal{F}\left[\psi(x,z)\right]\right], \tag{34}$$

where $\mathcal{F}$ denotes the discrete Fourier transform from z-space to p-space and $\mathcal{F}^{-1}$ denotes the inverse transform. Similarly, the WAPE operator is effected via

$$e^{i\Delta A/2}\psi(x,z) = \mathcal{F}^{-1}\left[e^{-i\frac{\Delta}{2}p^2/(k_0+\sqrt{k_0^2-p^2})}\mathcal{F}\left[\psi(x,z)\right]\right]. \tag{35}$$

Implementing the discrete Fourier transforms using FFT's leads to an efficient numerical algorithm—the split-step Fourier PE.

Both the SPE and WAPE approximations to the GPE propagator, $Q(x)$, produce errors in the field solution. To quantify them, consider the simplified problem of propagation in a vertically stratified atmosphere, in which case the wavenumber $K(x,z)$ takes the form Eq. (12). In the PE approximation, the field is $w^{PE}(x,z) = e^{ik_0 x}\psi(x,z)$, and ψ may be obtained by separation of variable methods. Let $\psi(x,z) = r(x)\phi(z)$ and substitute into Eq. (29). Then

$$\left(2ik_0\frac{\partial}{\partial x} + \lambda_m^2 - k_0^2 + \frac{1}{4x^2}\right) r(x) = 0, \tag{36}$$

and

$$\left(\frac{\partial^2}{\partial z^2} + k^2 m_{eff}^2(z) - \lambda_m^2\right)\phi(z) = 0, \tag{37}$$

where λ_m^2 is the separation constant. The discrete eigenfunction solutions, $\phi_m(z)$, of Eq. (37) correspond to propagating normal modes. The range equation (36) is easily solved and leads to

$$r_m(x+\Delta) = e^{i\frac{\Delta}{2k_0}\left(\lambda_m^2 - k_0^2 + \frac{1}{4x(x+\Delta)}\right)} r_m(x). \tag{38}$$

Thus a single mode in the PE approximation has a range dependence given by

$$w_m^{PE}(x+\Delta,z) = e^{i\frac{\Delta}{2k_0}\left(\lambda_m^2 + k_0^2 + \frac{1}{4x(x+\Delta)}\right)} w_m^{PE}(x,z). \tag{39}$$

If the same methods are applied to the full elliptic wave equation (10), with $w(x,z) = \hat{r}(x)\hat{\phi}(z)$, then the same z-seperated equation (37) is obtained for $\hat{\phi}(z)$, but Eq. (36) is replaced by

$$\left(\frac{\partial^2}{\partial x^2} + \lambda_m^2 + \frac{1}{4x^2}\right)\hat{r}(x) = 0. \tag{40}$$

The outgoing wave solution of Eq. (40) is just $\hat{r}_m(x) = \sqrt{x}\mathrm{H}_0^{(1)}(\lambda_m x)$. In the far-field, the exact elliptic modal field evolves with range according to

$$w_m(x+\Delta,z) = e^{i\Delta\left(\lambda_m + \frac{1}{8\lambda_m x(x+\Delta)}\right)} w_m(x,z), \tag{41}$$

where the asymptotic form of the Hankel function has been used.

Comparing equations (39) and (41) above, one sees that they are identical except for the phase. Thus in SPE, a normal mode will propagate with the correct amplitude and height-gain function, but with an incorrect phase velocity. For ranges larger than a few wavelengths, the range dependent terms in the exponents of equation (39) and (41) may be dropped, and yield the modal wavenumber in the PE approximation

$$\lambda_m' = \frac{\lambda_m^2 + k_0^2}{2k_0}. \tag{42}$$

Note that if the reference wavenumber $k_0 = \lambda_m$, then the parabolic and modal phase velocities are equal for the m-th mode and it will propagate correctly. In general, many modes contribute to the field and such a choice for k_0 is not possible.

There is, however, a way by which the PE reference wavenumber may be selected for multi-mode propagation which mitigates the PE phase error. As is known, an integral equation for the modal wavenumber, λ_m, may be written as[11]

$$\lambda_m^2 = \int \left[k^2 m_{eff}^2(z)\hat{\phi}_m^2(z) - \left(\frac{\partial\hat{\phi}_m}{\partial z}\right)^2\right] dz \Big/ \int \hat{\phi}_m^2(z)\,dz. \tag{43}$$

If the above equation is then averaged over all the propagating modes, an average modal wavenumber, $\bar{\lambda}$, results—this is the proper choice for the PE reference wavenumber:

$$k_C \equiv \bar{\lambda} = \int \left[k^2 m_{eff}^2(z)\psi^2(z) - \left(\frac{\partial\psi}{\partial z}\right)^2\right] dz \Big/ \int \psi^2(z)\,dz. \tag{44}$$

If this choice is made for k_0, then the reference wavenumber may be evaluated at each PE range step to reduce the parabolic approximation phase errors. This leads to a range dependent envelope transform similar to Eq. (13)

$$w(x,z) = e^{\int^x k_0(x')\,dx'}\psi(x,z), \tag{45}$$

with $k_0(x)$ evaluated by Eq. (44).

Magnus Expansion

Given the parabolic equation form Eq. (17), there still exists a problem of developing a numerical solution. The difficulty arises from the non-local nature of the Q operator. To solve Eq. (17) techniques from time-dependent quantum scattering theory will be used[12]. Define an evolution operator $U(\vec{x},\vec{x}_0)$ that determines the wave field $w(\vec{x})$ at the point $\vec{x}=(x,z)$ in terms of the wave field $w(\vec{x}_0)$ at the point $\vec{x}_0=(x_0,z)$:

$$w(\vec{x}) = U(\vec{x},\vec{x}_0)w(\vec{x}_0). \tag{46}$$

It follows by substitution into Eq. (17) that U is an operator satisfying the differential equation

$$\frac{\partial U(\vec{x},\vec{x}_0)}{\partial x} = iQ(\vec{x})U(\vec{x},\vec{x}_0). \tag{47}$$

Clearly, U must satisfy the initial condition

$$U(\vec{x}_0,\vec{x}_0) = 1, \tag{48}$$

and possess the group property

$$U(\vec{x},\vec{x}_0) = U(\vec{x},\vec{x}_1)U(\vec{x}_1,\vec{x}_0). \tag{49}$$

Futhermore, if there are no dissipative processes present, then it follows from the complex Poynting theorem that the z-integrated field energy density is a constant:

$$\int |w(x,z)|^2\,dz = \int |w(x_0,z)|^2\,dz. \tag{50}$$

Employing Eq. (46) this leads to the important result that the evolution operator U must be unitary:

$$U^{\dagger}(\vec{x},\vec{x}_0) = U^{-1}(\vec{x},\vec{x}_0), \tag{51}$$

where $U^{\dagger}$ is the (Hermitian) adjoint of U. The importance of having a unitary operator for numerical work should not be underestimated, since this form for U leads to numerical methods which can be shown to be absolutely convergent for all range step sizes.

Integrating Eq. (47) and using Eq. (48) yields the integral equation

$$U(\vec{x},\vec{x}_0) = 1 + i\int_{x_0}^{x} Q(x',z)U(x',x_0,z)\,dx'. \tag{52}$$

If Q were a c-number function instead of an operator, then Eq. (52) would be a Volterra integral equation since the independent variable x is the upper limit of integration. The theory of Volterra integral equations is well understood, and a Neumann series solution is easily found by iteration. If the same iterative method is applied to Eq. (52) then Dyson's expansion for U results:[12]

$$\begin{aligned} U(x,x_0,z) = 1 &+ i\int_{x_0}^{x} Q(x',z)\,dx' + (i)^2\int_{x_0}^{x} dx_1 \int_{x_0}^{x_1} dx_2\, Q(x_1,z)Q(x_2,z) \\ &+ (i)^3\int_{x_0}^{x} dx_1 \int_{x_0}^{x_1} dx_2 \int_{x_0}^{x_2} dx_3\, Q(x_1,z)Q(x_2,z)Q(x_3,z) + \cdots \end{aligned} \tag{53}$$

The ordering of the integrands in Eq. (53) is very important because, in general, the Q operators at different positions do not commute. Thus the ordering of the operators is determined by the range to which they refer, with operators refering to earlier ranges always appearing to the right. Though the Dyson expansion is a formally exact solution for U, if it is truncated after a finite number of terms, the result is no longer a unitary operator.

A solution which preserves the unitary character of the evolution operator U can be found by using a technique developed by Magnus[13]. To develop the Magnus expansion, introduce an iteration parameter λ into Eq. (47), and let $y(\lambda,x)$ satisfy

$$\frac{\partial y(\lambda,x)}{\partial x} = i\lambda Q(x)y(\lambda,x), \qquad y(\lambda,x_0) = 1, \tag{54}$$

and seek to join $y(0,x)=1$ to $y(1,x)=U(x,x_0)$. Assume that a solution of Eq. (54) is of the form

$$y(\lambda,x) = e^{\Omega(\lambda,x)}, \tag{55}$$

where

$$\Omega(\lambda,x) = \sum_{n=1}^{\infty} \lambda^n \Delta_n(x). \tag{56}$$

We now make use of the operator identity[14]

$$\frac{\partial}{\partial\lambda}e^{A(\lambda)} = \int_0^1 e^{sA(\lambda)}\frac{\partial A}{\partial\lambda}e^{-sA(\lambda)}\,ds\,e^{A(\lambda)}, \tag{57}$$

and obtain in view of equations (54) and (55)

$$\int_0^1 e^{s\Omega(\lambda,x)}\frac{\partial\Omega}{\partial x}e^{-s\Omega(\lambda,x)}\,ds = i\lambda Q(x). \tag{58}$$

Next, make use of the well-known commutator expansion[15]

$$e^A B e^{-A} = B + [A,B] + \frac{1}{2}[A,[A,B]] + \frac{1}{3!}[A,[A,[A,B]]] + \cdots \tag{59}$$

Substitute Eq. (56) for $\Omega(\lambda,x)$ and integrate over s (note: $\int_0^1 s^j\,ds = 1/j+1$) to obtain

$$\left\{\sum_{k=0}^{\infty}\frac{\Omega^k}{(k+1)!},\left(\sum_{n=1}^{\infty}\lambda^n\dot{\Delta}_n(x)\right)\right\} = i\lambda Q(x), \qquad \text{where} \quad \dot{\Delta}_n(x) = \frac{\partial\Delta_n(x)}{\partial x}. \tag{60}$$

The curlv brackets denote the iterated commutator which is defined inductively as

$$\left\{A^0,B\right\} = B, \qquad \left\{A^{n+1},B\right\} = [A,\{A^n,B\}]. \tag{61}$$

Equating the coefficients of λ^j on both sides of Eq. (60) yields recursive equations for the $\Delta_1,\Delta_2,\ldots$, etc. For $j=1$,

$$\dot{\Delta}_1(x) = iQ(x), \qquad \text{with} \quad \Delta_1(x) = i\int_{x_0}^{x}Q(x')\,dx'. \tag{62}$$

For $j=2$, one obtains

$$2\dot{\Delta}_2(x) = -\left[\Delta_1(x),\dot{\Delta}_1(x)\right]. \tag{63}$$

Hence

$$\Delta_2(x) = -\frac{1}{2}\int_{x_0}^{x}dx_1\int_{x_0}^{x_1}dx_2\,\left[Q(x_1),Q(x_2)\right]. \tag{64}$$

Similarly, for $j=3$

$$\Delta_3(x) = -\frac{i}{6}\int_{x_0}^{x}dx_1\int_{x_0}^{x_1}dx_2\int_{x_0}^{x_2}dx_3\,\left([[Q(x_1),Q(x_2)],Q(x_3)] + [[Q(x_3),Q(x_2)],Q(x_1)]\right). \tag{65}$$

Letting $\lambda \to 1$, we obtain the Magnus expansion for U:

$$\begin{aligned} U(x,x_0) = \exp\Bigg\{ & i\int_{x_0}^{x}Q(x')\,dx' - \frac{1}{2}\int_{x_0}^{x}dx_1\int_{x_0}^{x_1}dx_2\,\left[Q(x_1),Q(x_2)\right] \\ & -\frac{i}{6}\int_{x_0}^{x}dx_1\int_{x_0}^{x_1}dx_2\int_{x_0}^{x_2}dx_3\,\left([[Q(x_1),Q(x_2)],Q(x_3)] + [[Q(x_3),Q(x_2)],Q(x_1)]\right) + \cdots\Bigg\} \end{aligned} \tag{66}$$

Split-Step Errors

While the Magnus expansion provides an exact solution to the PE wave field, three approximations must be made to it before numerical computations are feasible. First, the infinite series in the Magnus exponent is truncated after the first term

$$U(x_0+\Delta,x_0) \approx e^{i\Delta\bar{Q}}, \qquad \bar{Q} = \frac{1}{\Delta}\int_{x_0}^{x_0+\Delta}Q(x')\,dx', \tag{67}$$

where $\bar{Q}$ is the range-averaged PE propagator. Second, the pseudo-differential PE propagator operator, Q, is approximated by the sum of two non-commuting operators (for example, the for n specified by equation (27) or (28))

$$Q(x,z) \approx A(z) + B(x,z). \tag{68}$$

Finally, the exponential Magnus solution is factored (split) in the symmetrized form

$$e^{i\Delta(A+\bar{B})} \approx e^{i\Delta A/2}\,e^{i\Delta\bar{B}}\,e^{i\Delta A/2}; \qquad \bar{B} = \frac{1}{\Delta}\int_{x_0}^{x_0+\Delta}B(x',z)\,dx'. \tag{69}$$

The truncation errors incurred by use of Eq. (67) are a function of the range dependence of Q. If Q, or equivalently K, is independent of x, then Eq. (67) is exact. For range dependent K, the truncation error in U may be quantified by examining the next term in the Magnus expansion and gives

$$\begin{aligned} U(x_0+\Delta, x_0) &\approx \exp\left\{ i\Delta\bar{Q} - \frac{1}{2}\int_{x_0}^{x_0+\Delta} dx_1 \int_{x_0}^{x_1} dx_2 [Q(x_1), Q(x_2)] \right\}, \\ &= \exp\left\{ i\Delta\bar{Q} - \frac{\Delta^3}{6}\left[\frac{\partial Q}{\partial x}, Q\right] + O(\Delta^4)\right\}. \end{aligned} \tag{70}$$

Thus, when Q is range dependent, truncation errors proportional to $O(\Delta^3)$ are incurred no matter what form of the PE propagator is used. Furthermore, how the range-averaged PE propagator $\bar{Q}$ is evaluated critically effects the error budget. For example, if Q is expanded in a Taylor series about the end-point x_0, then the evolution operator becomes

$$U(x_0+\Delta, x_0) = \exp i\left\{ \Delta Q(x_0) + \frac{1}{2}\Delta^2 Q'(x_0) + \frac{1}{6}\Delta^3\left(Q''(x_0) + i[Q'(x_0), Q(x_0)]\right) + O(\Delta^4)\right\}, \tag{71}$$

where $Q'(x_0) = \left.\frac{\partial Q(x)}{\partial x}\right|_{x=x_0}$, and $Q''(x_0) = \left.\frac{\partial^2 Q(x)}{\partial x^2}\right|_{x=x_0}$; while if $\bar{Q}$ is expanded about the mid-point $\bar{x} = (x+x_0)/2$ then

$$U(x_0+\Delta, x_0) = \exp\left\{ i\Delta Q(\bar{x}) + \frac{\Delta^3}{24}(iQ''(\bar{x}) - 4[Q'(\bar{x}), Q(\bar{x})]) + O(\Delta^4)\right\}. \tag{72}$$

Thus the truncation error, E_1, incurred in the field solution by using Eq. (67) with Q evaluated at the mid-point x_0 is

$$\begin{aligned} E_1 &\equiv \left[U(x_0+\Delta, x_0) - e^{i\Delta Q(\bar{x})}\right]\psi(x_0), \\ &\approx \frac{\Delta^3}{24}\left(iQ''(\bar{x}) - 4[Q'(\bar{x})Q(\bar{x}) - Q(\bar{x})Q'(\bar{x})]\right)\psi(x_0). \end{aligned} \tag{73}$$

A second type of truncation error occurs when the splitting approximation, Eq. (69), is employed. This splitting error can be evaluated by using the Baker-Campbell-Hausdorff[14-15] (BCH) expansion of two non-commuting operators:

$$e^A e^B = \exp\left(A + B + \frac{1}{2}[A,B] + \frac{1}{12}[A,[A,B]] + \frac{1}{12}[[A,B],B] + \cdots\right). \tag{74}$$

Applying the BCH expansion to Eq. (69) gives

$$e^{i\Delta A/2}e^{i\Delta\bar{B}}e^{i\Delta A/2} = \exp i\left(\Delta A + \Delta\bar{B} - \frac{\Delta^3}{12}[[A,\bar{B}],\bar{B}] + \frac{\Delta^3}{24}[[B,A],A] + O(\Delta^4)\right), \tag{75}$$

and the local truncation error, E_2, in the PE solution with B evaluated at the mid-point is thus

$$\begin{aligned} E_2 &\equiv \left[e^{i\Delta(A+B(\bar{x}))} - e^{i\Delta A/2}e^{i\Delta\bar{B}}e^{i\Delta A/2}\right]\psi(x_0) \\ &= \frac{i}{24}\Delta^3\left\{2[[A,B(\bar{x})],B(\bar{x})] - [[B(\bar{x}),A],A]\right\}\psi(x_0) + O(\Delta^4). \end{aligned} \tag{76}$$

If the standard PE propagator approximation, Eq. (27), is used, then the local truncation error term is given explicitly by

$$\begin{aligned} \delta\psi(x_0+\Delta, z) = \frac{i\Delta^3}{48k_0}\Bigg\{ &\left[\frac{1}{4}\frac{\partial^4 V(\bar{x})}{\partial z^4} + ik_0\frac{\partial^2 V'(\bar{x})}{\partial z^2} - k_0^2\left(\frac{\partial^2 V(\bar{x})}{\partial z^2}\right)^2\right]\psi(x_0) \\ &+ \left[i2k_0\frac{\partial V'(\bar{x})}{\partial z} - \frac{\partial^3 V(\bar{x})}{\partial z^3}\right]\frac{\partial\psi(x_0)}{\partial z} + \frac{\partial^2 V(\bar{x})}{\partial z^2}\frac{\partial^2\psi(x_0)}{\partial z^2}\Bigg\}, \end{aligned} \tag{77}$$

where $V(\bar{x}) = 1 - K^2(\bar{x}, z)/k_0^2$, and $V'(\bar{x}) = \left.\frac{\partial V(x)}{\partial x}\right|_{x=\bar{x}}$. Examination of Eq. (66) shows that the local truncation error for the symmetrized splitting is linear in frequency (via the k_0 terms) and cubic in the range step-size Δ. A similar error analysis results if the wide-angle PE propagator, Eq. (28), is used; the form is identical with Eq. (77) except that $V(\bar{x}, z) = K(\bar{x}, z) - k_0$.

In numerical implementation, the vertical derivatives appearing in Eq. (77) are evaluated by numerical finite difference approximations. As the RPE code advances the field, the local error budget is monitored and the range step-size is dynamically adjusted to keep the local error below a threshold.

4. CONCLUSION

This paper has focused on some of the issues relating to error control in parabolic wave equation codes and how to implement them numerically. In the course of developing the RPE code, much effort was devoted to user interfaces and automated techniques to ensure the fidelity of the outputs. The RPE code incorporates the error checking features described earlier and implements the split-step Fourier PE algorithm. The code is written in FORTRAN and currently operates on a wide spectrum of computers ranging from INTEL 80286/80386 personal computers , SUN mini-computers, and VAX mainframes. Work is currently underway to adapt RPE to PC-based SKY VORTEX array processors and to a VME-bus systolic processor. This will lead to substantial decreases in CPU resources.

ACKNOWLEDGMENTS

This work was supported in part by the Office of Naval Technology and in part by the Naval Ocean Systems Center Independent Research program.

REFERENCES

1. M. A. Leontovich and V. A. Fock, *Solution of the problem of propagation of electromagnetic waves along the earth's surface by the method of parabolic equations*, J. Phys. of the USSR **10** (1946), 13-24.
2. V. A. Fock, "Electromagnetic Diffraction and Propagation Problems," Chap. 11-14, Pergamon Press, 1965.
3. R. H. Hardin and F. D. Tappert, *Analysis, simulation and models of ionospheric scintillation*, Bell Labs Report (1974).
4. R. H. Hardin and F. D. Tappert, *Application of the split-step Fourier method to the numerical solution of nonlinear and variable coefficient wave equations*, SIAM Rev. **15** (1973), p. 423.
5. H. W. Ko, J. W. Sari, and J. P. Skura, *Anomalous microwave propagation through atmospheric ducts*, Johns Hopkins APL Tech. Dig. **4** (1983), 12-16.
6. C. L. Pekeris, *Accuracy of the earth-flattening approximation in the theory of microwave propagation*, Phys. Rev. **70** (1946), 518-522.
7. F. D. Tappert, *The parabolic approximation method*, in "Wave Propagation and Underwater Acoustics," Lecture Notes in Physics, Ed. by J. B. Keller and J. S. Papadakis, Springer-Verlag, 1977, pp. 224-287.
8. H. Bremmer, *The WKB approximation as the first term of a geometric optical series*, in "Theory of Electromagnetic Waves Symposium," Interscience Publishers, Inc., New York, 1951, p. 169.
9. J. Corones, *Bremmer series that correct parabolic approximations*, J. Math. Anal. Appl. **50** (1975), 361-372.
10. M. D. Feit and J. A. Fleck, Jr., *Light propagation in graded-index fibers*, Appl. Opt. **17** (1978), 3990-3998.
11. A. W. Snyder and J. D. Love, "Optical Waveguide Theory," Chapman and Hall, Ltd., 1983, p. 643.
12. C. Joachain, "Quantum Collision Theory," Third Ed., North Holland, 1983.
13. W. Magnus, *On the exponential solution of differential equations for a linear operator*, Comm. Pure and Appl. Math. **7** (1954), 649-673.
14. R. M. Wilcox, *Exponential operators and parameter differentiation in quantum physics*, J. Math. Phys. **8** (1967), 962-982.
15. E. Merzbacher, "Quantum Mechanics," John Wiley and Sons, Inc., 1961, p. 162.
16. H. F. Baker, *On the integration of linear differential equations*, Proc. London Math. Soc. **34** (1902), p. 347;, **35** (1903), p. 333;, **Series 2** (1904), p. 293.
17. J. E. Campbell, Proc. London Math. Soc. **29** (1898), p. 14.
18. F. Hausdorff, Ber. Verhandl. Saechs. Akad. Wiss. Leipzig, Math. Naturw. Kl. **58** (1906), p. 19.

A FORECASTING SYSTEM USING THE PARABOLIC EQUATION — APPLICATION TO SURFACE-TO-AIR PROPAGATION IN THE PRESENCE OF ELEVATED LAYERS

by

K.H. Craig and M.F. Levy
Radio Communications Research Unit
Rutherford Appleton Laboratory
Chilton, Didcot, OX11 0QX
United Kingdom

SUMMARY

The parabolic equation approach to clear-air propagation modelling overcomes many of the difficulties associated with ray and mode theory methods. A parabolic equation model has been implemented on a PC based system using a transputer to carry out the computationally intensive numerical integrations. The model has been used from VHF to millimetric frequencies and applied to evaporation duct and elevated duct problems. The latter are important for surface-to-air propagation and have been difficult to solve because of the complicated structure of the layers. A case study of an elevated duct caused by anticyclonic subsidence shows the importance of up-to-date meteorological data from a wide geographical area. A full-wave calculation of the wideband properties of the propagation channel illustrates the possibilities opened up by the new model. The frequency selective effects can be large, and are sensitive to the small-scale structure of the ducting layers.

1. INTRODUCTION

Current forecasting aids for microwave propagation in the troposphere (such as IREPS [1]) were developed principally for the naval environment where the evaporation duct dominates propagation conditions. They are based on a combination of ray tracing and mode theory and assume a simple one-dimensional refractive index profile. For airborne applications, elevated layers in the troposphere assume a greater significance in the forecasting of radar coverage, and it is no longer adequate to assume horizontal homogeneity of the layers. Real-time mode theory calculations are prohibitively expensive and extending the "template matching" approach of IREPS to two-dimensional structures could be very cumbersome since a single mode approximation is not valid for elevated layers.

The rapid advances in computer capabilities over the last decade have led to a re-evaluation of purely numerical, but more general, solutions of the propagation problem as alternatives to the semi-analytical ray and mode methods. The best example of this is the parabolic equation approach to the solution of Maxwell's equations; this is being followed by several groups at this meeting. The theory underlying the parabolic equation [2,3,4] will not be repeated here. The method is computationally efficient due to a Fast Fourier Transform (FFT) implementation of the "split step" algorithm that was first applied to an analogous problem in underwater acoustics. The vertical field pattern at a range x is advanced to range $x + \Delta x$ by transforming the field (in z-space) into its angular spectrum (p-space, where $p = k \sin\theta$, k is the wavenumber and θ is the angle of propagation to the horizontal), applying an operator, transforming back to z-space and applying a further (refractive index dependent) operator. The parabolic equation has been found to be particularly effective for propagation through two-dimensional elevated structures and generates field strength/radar coverage diagrams more easily than either ray or mode methods.

The features and capabilities of a parabolic equation method propagation forecasting model implemented on a desk-top computer are described. Applications of the model to propagation in the evaporation duct and in the presence of inhomogeneous elevated ducting layers are presented. The latter show the importance of accounting for horizontal variations. The potential of the new model is illustrated by a calculation of the wideband properties of the propagation medium.

2. A PC-BASED PROPAGATION FORECASTING MODEL

A stand-alone desk-top propagation forecasting system is of great assistance to the engineer or meteorologist. The rapid advances in microprocessor technology and the falling prices of high-quality output devices have eliminated the need for mainframe resources for all but the most computationally intensive tasks. In the operational environment, mainframe systems are ruled out in any case. IREPS [1] is a good example of a user friendly desk-top forecasting system based on the technology available in the late 1970s. The authors investigated the possibility of implementing full mode theory calculations on a microcomputer, but it was clear that even for one-dimensional profiles, execution times would be unrealistically long at microwave and millimetre wave frequencies. The development of the parabolic equation method has resulted in a prediction program, PCPEM, that runs on an IBM PC-AT (or similar) based system. Some of the features of the model are:

- colour two-dimensional field strength diagrams are produced directly on screen; field strength/path loss data can be saved for off-line display or calculation of path loss versus range or height graphs;
- two-dimensional refractive index structures can be modelled analytically or numerically;
- surface conductivity and surface roughness effects on antenna lobing are included. Since the split-step algorithm alternates between z-space and p-space, the boundary conditions can be imposed through an approximate image theory approach: Fresnel coefficients are applied to each component of the angular spectrum of the image field, modified to include the usual roughness factor;
- gaseous absorption effects are modelled: the absorption at a particular frequency can be calculated at each point from the pressure, temperature and humidity. This becomes important at millimetre waves where the large absorption coefficients and rapidly varying humidity within ducting structures can give rise to complex field patterns.

Computation times increase with frequency and with larger range and height requirements, as both these factors require larger FFT sizes in order to satisfy the Nyquist criterion in z or p-space. It was not considered practical to run the model on a PC by itself since run times could be up to an hour for short range (50 km) evaporation duct problems. The solution was to increase the power of the system by using an add-in processor card on which the bulk of the numerical processing takes place, based on the INMOS T-800 floating point transputer [5]. The transputer is a 32-bit CMOS reduced instruction set (RISC) chip capable of 1.5 Mflops (20 MHz clock), giving a speed increase of 40–50 times over a 10 MHz 80286 based PC. The transputer has further potential advantages due to its unique architecture: it was specifically designed to allow easy implementation of parallel processing on multi-transputer networks. As well as having separate integer and floating point units, the chip contains all the necessary clock and memory circuitry, a hardware task scheduler, and four independent fast (20Mbit/s) asynchronous serial links. A single transputer system is cheap and easy to design, as very little "glue" circuitry is required, and a multi-processor system can be built up by connecting the single processors together with single wire links, as all the necessary hardware synchronisation protocols are contained on-chip.

A feature of the transputer architecture is that the instruction set has been designed as an efficient target of high level languages. Fortran 77, C and Pascal are all available, and compiled code can yield a sustained performance of around 1 Mflops on a single transputer. This translates into times of 0.1 and 2 seconds for compiled double precision complex FFTs of sizes 1k and 16k respectively. A basic single transputer card with 2 Mbytes of memory allows complex FFT sizes up to 64k and will solve a typical evaporation duct problem (50 km by 100 m frame) in 1–2 minutes at 10 GHz or 5 minutes at 100 GHz; a long range elevated duct problem (*e.g.* 200 km by 1 km frame) takes about 40 minutes at 10 GHz. Commercial add-in processor cards are available for the PC with up to seventeen transputers on a single card. The FFT algorithm that is at the heart of PCPEM is an ideal candidate for parallel processing, and the authors are currently developing such a version. As an example, a 64 transputer array reduces execution times by a factor of 40.

3. PROPAGATION IN THE EVAPORATION DUCT

Forecasting of propagation in the evaporation duct has been a fruitful area of application of mode theory methods. Their success owes much to the assumption that the evaporation duct is horizontally uniform. There have, however, been two main drawbacks in these approaches. Firstly, as stated above, computational times are long, particularly at the higher microwave frequencies, and operational systems such as IREPS rely on the use of "template matching" whereby precalculated mode theory results are scaled to the frequency and duct height of interest. This can be successful at lower frequencies where single mode propagation dominates, but would become cumbersome at the higher frequencies when multi-mode propagation is important. A second drawback of mode theory is the difficulty of producing coverage diagrams: mode theory does not apply in the line-of-sight region close to the antenna and a hybrid ray and mode method has to be used.

The parabolic equation produces full-wave two-dimensional coverage diagrams directly. Figure 1 is a series of diagrams showing the effects of the evaporation duct on a horizontally polarised antenna at 10 m. More precisely, the diagrams show the one-way basic transmission loss. To illustrate the effects of antenna lobing due to surface reflections, a smooth sea has been assumed. The diagrams give results at S-band and X-band for duct heights ranging from 0 m (pure diffraction) to 30 m. At S-band, the effect is one of increasing detection range beyond the horizon with increasing duct height. At X-band, the higher duct heights show a more complicated pattern of coverage holes; the appearance of "islands" in the coverage pattern would not be apparent from simple ray tracing where the contours are always connected. Likewise, many modes would be required to describe this structure in the mode theory formalism. At higher frequencies, these effects become more pronounced, particularly when gaseous absorption is included at millimetric wavelengths [6].

4. SURFACE-TO-AIR PROPAGATION IN THE NORTH SEA REGION

The main reason for the development of the model described here was an interest in the effects of elevated tropospheric layers on transhorizon propagation. One application was the problem of coordination of civil terrestrial and satellite communication systems within Europe, and is discussed in [6]. A second was the forecasting of the propagation effects on surface-to-air radar and communications systems in the North Sea region between the U.K. and Northern Europe.

The main problem for propagation forecasting in the presence of elevated layers is the difficulty of obtaining sufficient meteorological data in the first place. Although new remote sounding systems are being introduced, the 12 hourly meteorological radiosonde ascents are still the most widely available sources of data. In order to assess the adequacy of radiosonde coverage in the North Sea region, data from the midnight and midday ascents at thirteen stations have been obtained during periods of strong anticyclonic activity. Despite the poor height resolution of the significant level data, 118 of the 290 ascents available (41%) showed the presence of at least one ducting layer. The overwhelming majority of these were elevated ducts. There were quite significant spatial and temporal variations from ascent to ascent. A detailed examination of one of the events is now given.

During the period of 4–7th November 1987, a major anticyclone was centred over Northern Europe, moving slowly south-east and declining. Figure 2 shows a map of the southern North Sea region with the radiosonde stations at Aughton, Hemsby, de Bilt and Essen shown. These were selected as they lie very nearly on a straight line from England, across the southern North Sea, into Europe. In order to assess the effect of the anticyclonic subsidence on propagation, it was assumed that a 3 GHz airborne emitter was located near the inversion layer at a height of 750 m over Essen. Figure 3 shows the coverage diagram for this source under *standard* atmospheric conditions. The locations of Essen, de Bilt and Hemsby are shown—note that the figure shows the path from east to west. A receiver sited at Hemsby on the U.K. coast (a distance of 388 km from Essen) lies well beyond the horizon (113 km).

The synoptic charts for 1200h on the 6th, and 0000h and 1200h on the 7th are shown in Figure 4, and the modified refractivity profiles for these three times measured at Aughton, Hemsby, de Bilt and Essen are shown in Figure 5. A superrefracting/ducting layer lies between 600 and 1000 m in altitude,

with a significant downward slope towards the U.K. The level of the inversion on the Continent falls during the 24 hour period as the centre of the anticyclone moves south-east. The development at the U.K. end is less straightforward, although there is some evidence of an upward trend at Aughton. A uniform, homogeneous layer is certainly not a good model in this case.

Interpolation of the two-dimensional refractivity data of Figure 5 was eased by the simple structure of the layers. The essential features of each ascent were extracted by hand, and each vertical profile was modelled by a simple *tanh* layer superimposed on an exponential atmosphere. At any range, the heights of the top and bottom of the layer, and the refractivity change across the layer, were found by linear interpolation between adjacent ascents. For more complex layer structures, a more sophisticated interpolation scheme based on *a priori* knowledge of the meteorological structures is required. Figure 6 shows the coverage diagrams for 1200h on the 6th, and 0000h and 1200h on the 7th November 1987 for the same emitter, using the modified refractivity profiles of Figure 5. The superrefracting layer over the Continent on the 6th is too weak and too high to have much effect on coverage. As the layer descends and strengthens during the evening, strong ducting occurs, giving rise to high fields over the U.K. By midday on the 7th, the layer has weakened again, although a considerably enhanced propagation path is still apparent. The need for up-to-date meteorological data for forecasting in the presence of elevated layers is obvious—both the subsidence and advection conditions that give rise to such layers are subject to large diurnal variations.

A ground based receiver in the U.K. would not be able to take advantage of the enhanced propagation in this case. Figure 7 shows the situation from the U.K. end of the path. A receiver has been placed at 50 m height at Hemsby (on the left). No signal will be received by a receiver located so far below the duct. Indeed, the duct has very little effect on the coverage diagram at all. However, since there is a high level of leakage above the duct, a receiver positioned anywhere in this region would be capable of detecting the emitter; it is not necessary for the receiver to be located within the duct.

Figure 8 shows the coverage diagram that would have been obtained at 0000h on the 7th November based on the Hemsby ascent *if it had been assumed that this ascent were representative of the whole path.* This assumption very substantially underestimates the signal levels detectable in the U.K. (by more than 20 dB) compared with the true situation of Figure 6b. Two-dimensional meteorological data, and a propagation model that uses it, are clearly necessary.

5. WIDEBAND EFFECTS

The effects of the propagation medium on high bit rate digital communications and on short pulse radar systems are complex because of the frequency selective nature of multipath and ducting. These effects are characterised by the transfer function of the propagation channel, which shows field strength relative to free space as a function of frequency. In the ray optics approach to the calculation of transfer functions, the amplitudes and optical path lengths of the multiple rays connecting transmitter and receiver are obtained and these are then added coherently. This method is difficult to automate, as it is difficult to identify all the relevant rays. (In ray tracing, a single ray does not carry any amplitude information. A pencil of rays is required, and these pencils can become severely distorted in a ducting environment, making it difficult to identify rays from different "families"). Ray optical calculations of field amplitudes also fail at the caustic surfaces formed close to the ducting layers. The parabolic equation method eliminates these problems by providing a full-wave solution. Two-dimensional range–height calculations of field strength relative to free space are made at a set of frequencies over the bandwidth of interest, and the channel transfer function is extracted directly. Both the magnitude and group delay are available, and the effects of space diversity are easy to evaluate.

As an illustration, Figure 9 shows the field of a 10 GHz antenna at a height of 25 m over the sea in the presence of a 40 N-unit homogeneous ducting layer at 300 m. The horizon in standard conditions would be at 15 km, and the trapping and leakage from the top of the duct are very noticeable. A horizontal cross-section at 250 m (marked (a)–(e)) is shown, and the transfer function at each of these points along the cross-section (separated by 40 km) was calculated every 50 MHz between 10 GHz and 11 GHz. The results are shown in Figure 10. (Position (c) is deep within a coverage "hole" and has

a transfer function that is 35–40 dB below free space.) The wideband distortions are large (greater than 10 dB) and change substantially as the relative positions of transmitter and receiver vary. This could have a significant effect on signature analysis for wideband radars, and would cause increased intersymbol interference on wideband communications links. It should be added that the periodicity observed in Figure 10 cannot be explained in any simple way by means of two- or three-ray interference models. A fuller discussion of wideband effects is given in [7].

The sensitivity of the wideband results to the relative position of transmitter and receiver has important implications. The wideband results above assumed a horizontally homogeneous layer. A more realistic situation is shown in Figure 11. The geometry is the same as in Figure 9, except that the layer is "corrugated", the corrugations having an amplitude of 25 m and a wavelength of 5 km, typical of the structures observed in stratocumulus layers at the top of subsidence inversions [8]. On the large scale the main effect has been a smoothing out of the field (filling in the coverage holes) and an increased leakage from the top of the duct, but the detailed field pattern is complicated (varying by up to 20 dB in the space of a few kilometres). The field is sensitive to the detailed structure of the layers, and the meteorology is unlikely to be measured in sufficient detail for completely deterministic forecasts. Rather, a statistical description of the field may be required [6].

6. CONCLUSIONS

The recent rapid advances in desk-top computer capabilities have led to the development of a forecasting model for a PC based system that will display the effects of anomalous propagation on communications and radar systems. Based on a full-wave parabolic equation solution, execution times are similar to earlier ray optics implementations. The model has been applied to short range evaporation duct and long range elevated duct propagation at frequencies as high as 100 GHz. Problems that were not easily solved by ray tracing (such as wideband effects) can now be tackled. For surface-to-air propagation, there is a need for up-to-date two-dimensional radiometeorological data if accurate forecasts are to be made. The main difficulty now is obtaining high quality data in the first place.

REFERENCES

[1] Baumgartner G.B., Hitney B.A. and Pappert R.A., "Duct propagation modelling for the integrated-refractive-effects prediction system (IREPS)", *Proc. IEE* **130**, Part F, 1983, 630–642.

[2] Fock V.A., "Electromagnetic diffraction and propagation problems", Pergamon Press, London, England, 1965, chapters 11,13,14.

[3] Tappert F.D., "The parabolic approximation method", in *Wave propagation and underwater acoustics*, Lecture Notes in Physics, **70**, Springer Verlag, Berlin, 1977, 224–287.

[4] Ko H.W., Sari J.W., Thomas M.E., Herchenroeder P.J. and Martone P.J., "Anomalous propagation and radar coverage through inhomogeneous atmospheres", *AGARD Conf. Proc.* **346**, 1984, 25.1–14.

[5] INMOS Limited, "Transputer reference manual", Prentice Hall, London, 1988.

[6] Levy M.F. and Craig K.H., "Assessment of anomalous propagation predictions using minisonde refractivity data and the parabolic equation method", 1989, this publication.

[7] Craig K.H. and Levy M.F., "Field strength forecasting with the parabolic equation: wideband applications", *Int. Conf. on Antennas and Propagation (ICAP 89)*, IEE Conf. Pub. **301**, 1989.

[8] Nicholls S. and Leighton J., "An observational study of the structure of stratiform cloud sheets: Part 1. Structure", *Quart. J. R. Met. Soc.* **112**, 1986, 431–460.

ACKNOWLEDGEMENTS

This work has been carried out in part with the support of the Procurement Executive, Ministry of Defence. The authors would like to thank J.D. Turton and S.F.G. Farmer of the U.K. Meteorological Office for several discussions and for providing the radiosonde data used in this pape

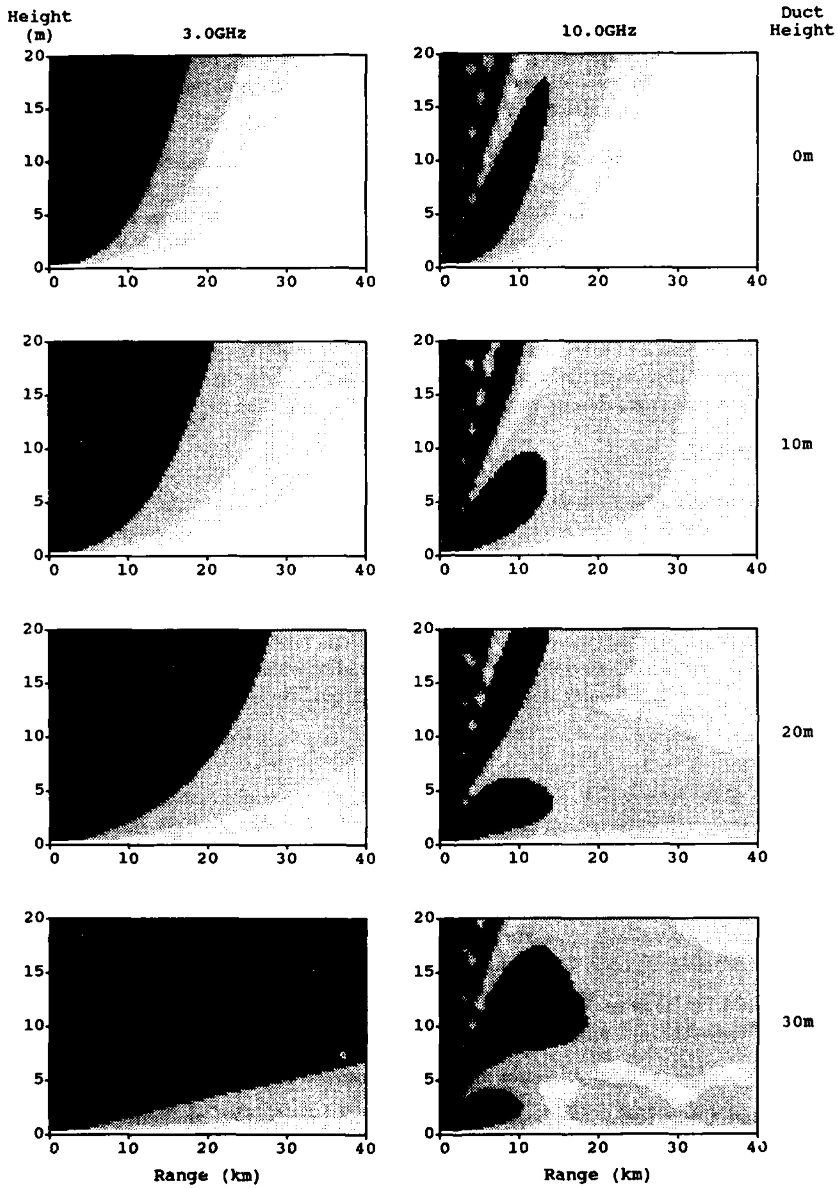


Figure 1: The diagrams show contours of basic transmission loss for a horizontally polarised antenna at 10 m over a smooth sea. The two columns are for 3 GHz and 10 GHz. Evaporation duct heights of 0 m (pure diffraction), 10 m, 20 m and 30 m are shown. (The contour spacing is 10 dB and the darkest level represents a transmission loss less than 110 dB).

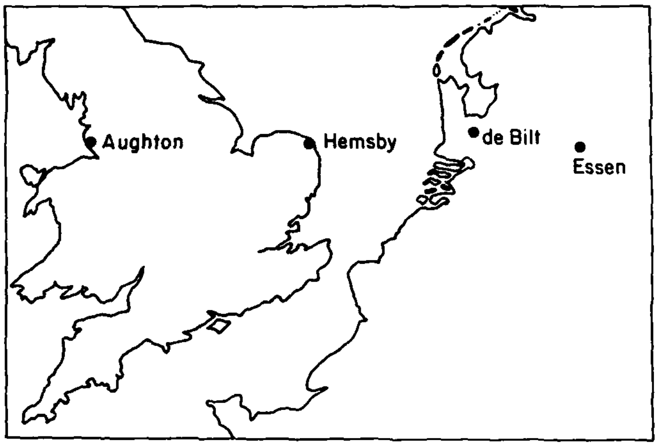


Figure 2: Map of the southern North Sea and English Channel showing the positions of the radiosonde stations.

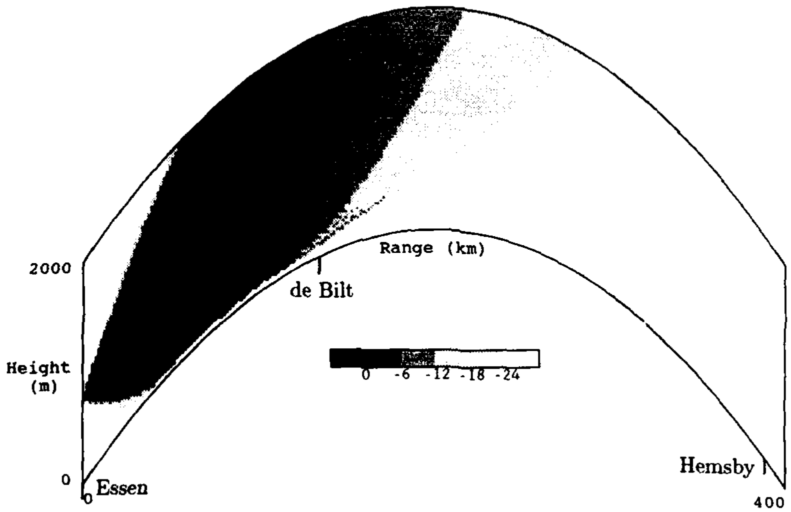


Figure 3: Coverage diagram of a 3 GHz airborne emitter at 750 m over Essen: standard atmospheric conditions. (Contours in dB).

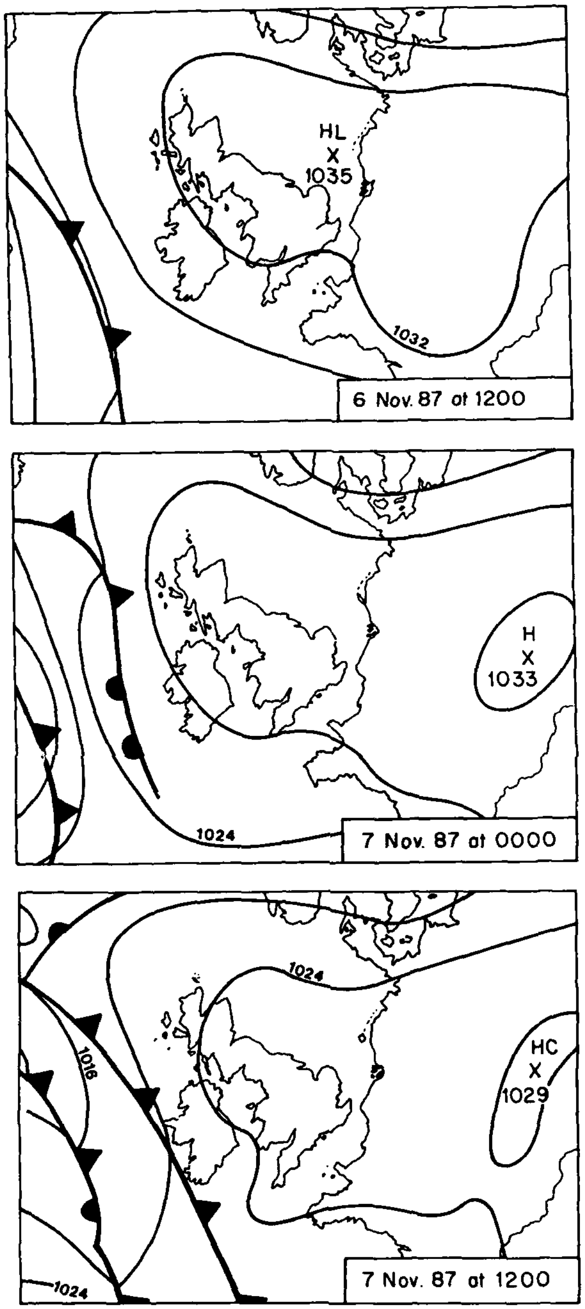


Figure 4: Synoptic charts for 6th and 7th November 1987.

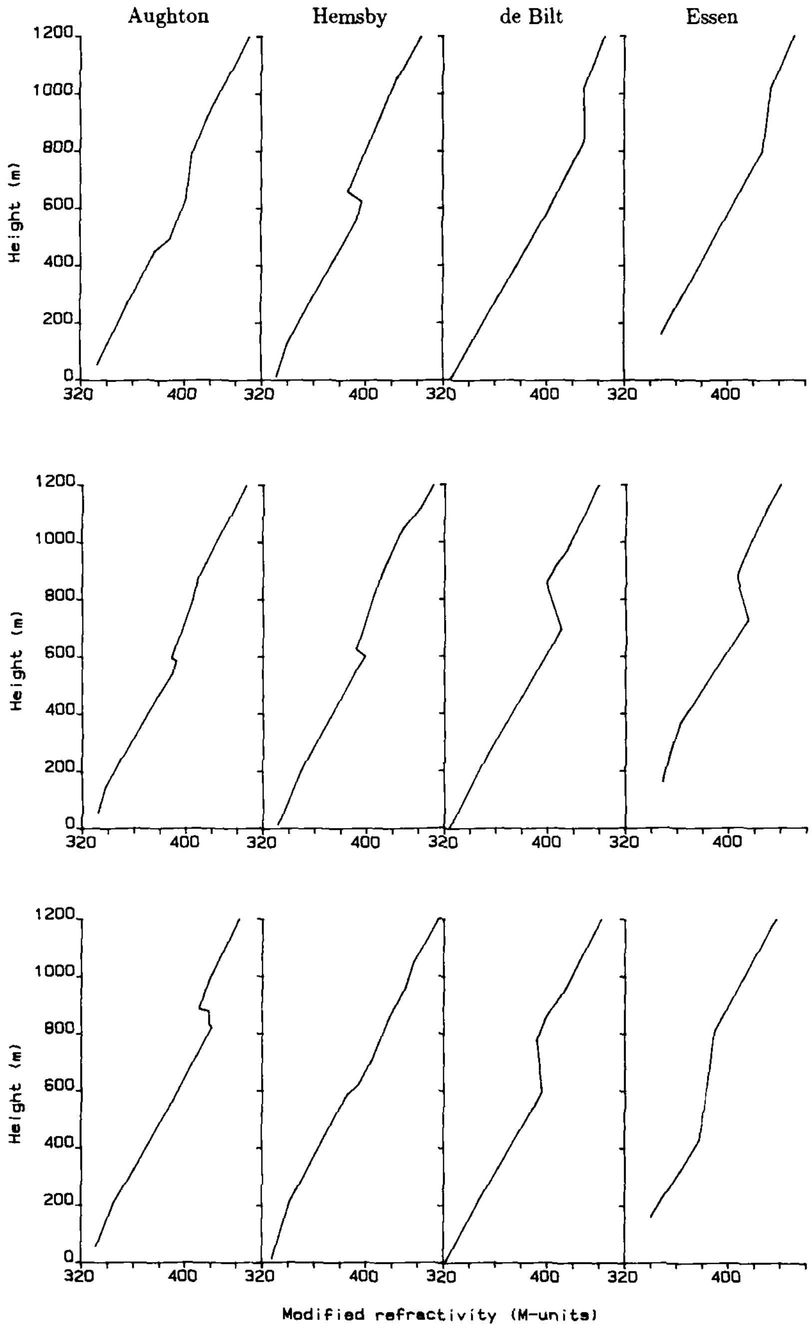


Figure 5: Modified refractivity profiles derived from the radiosonde ascents made at the time of the corresponding synoptic charts in Figure 4.

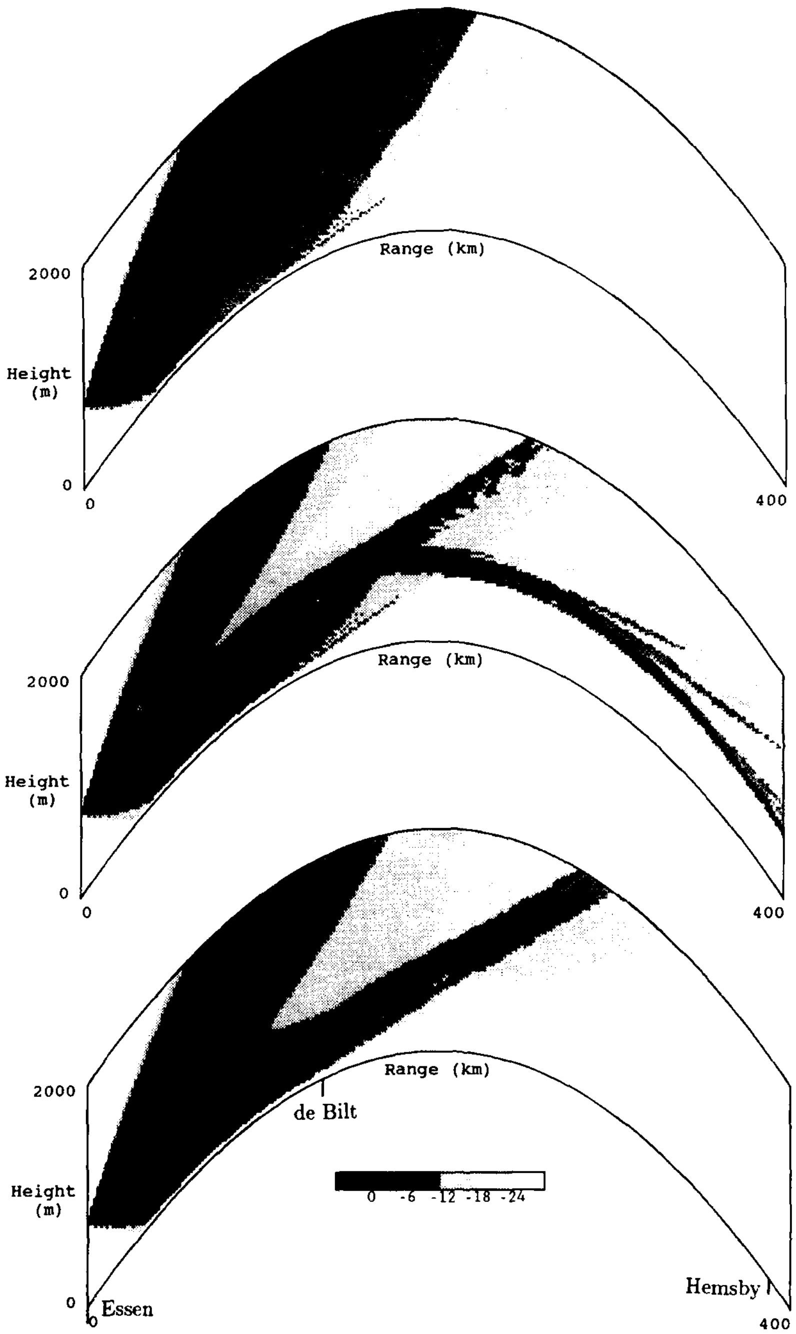


Figure 6: Coverage diagram of a 3 GHz airborne emitter at 750 m: the refractivity profiles of Figure 5 were used: (top to bottom) 6th (1200h), 7th (0000h), 7th (1200h), November 1987. (Contours in dB).

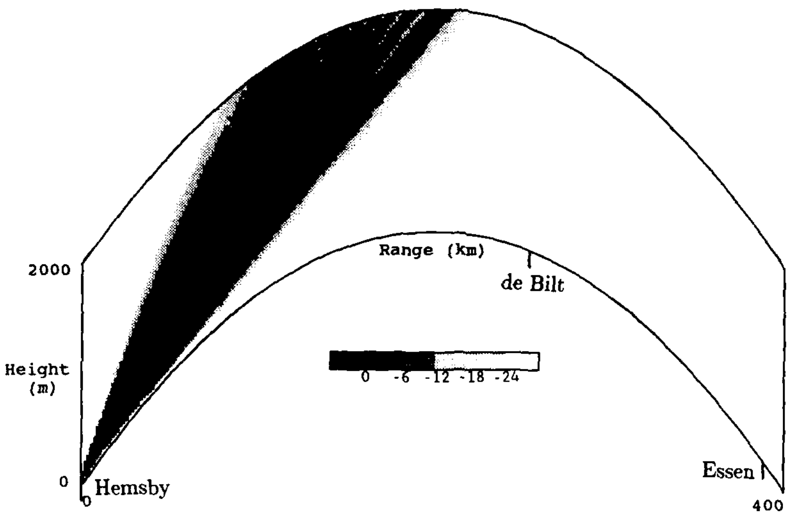


Figure 7: Coverage diagram of a 3 GHz sensor at 50 m located at Hemsby. Radiosonde data for 7th November (0000h). (Contours in dB).

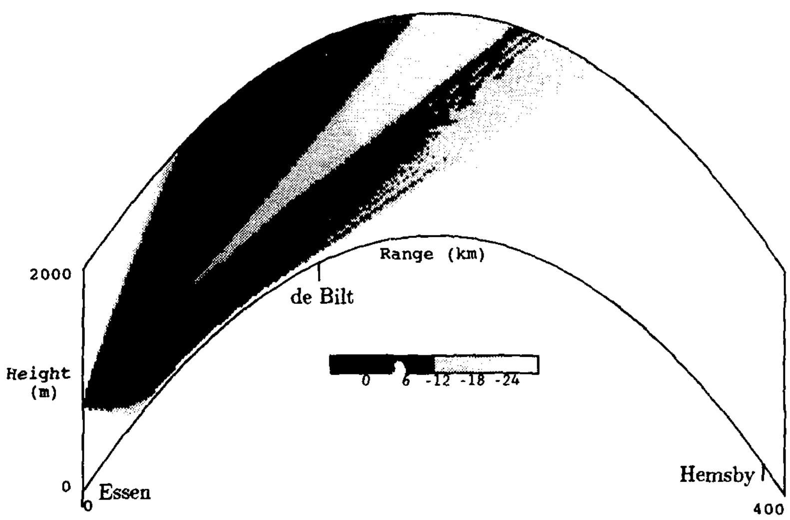


Figure 8: Coverage diagram of a 3 GHz airborne emitter at 750 m: 7th November (0000h)—assuming Hemsby data representative of whole path. (Contours in dB).

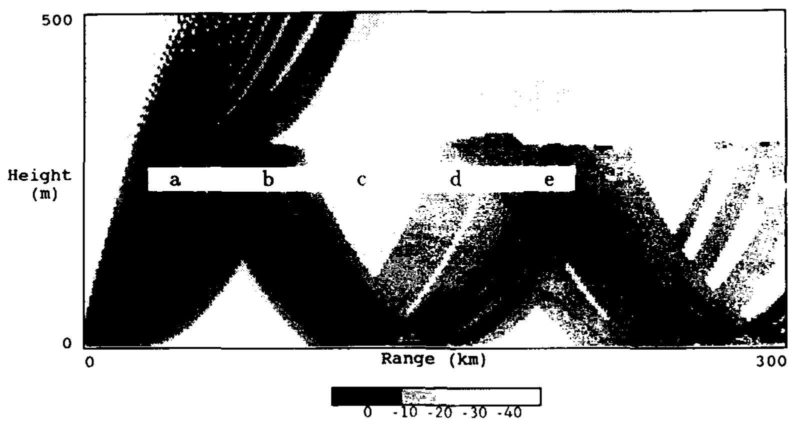


Figure 9: Field strength contours (dB) for a 10 GHz gaussian antenna at a height of 25 m over the sea. A 40 N-unit ducting layer is present at 300 m.

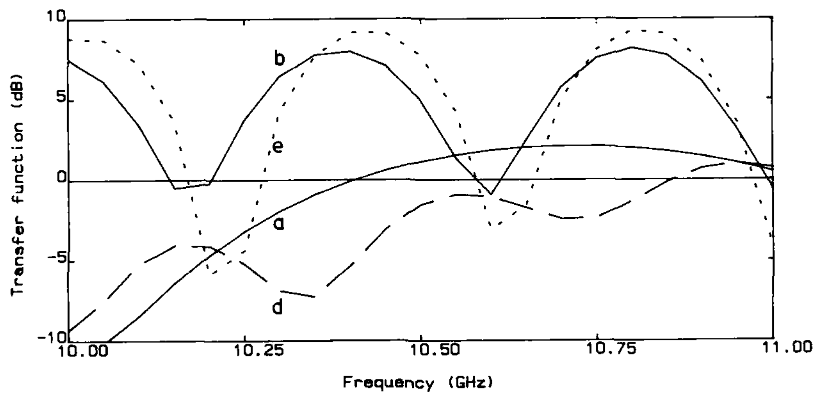


Figure 10: Channel transfer functions for positions (a), (b), (d) and (e) of Figure 9. (Position (c) has a transfer function that is 35–40 dB below free space and is not shown.)

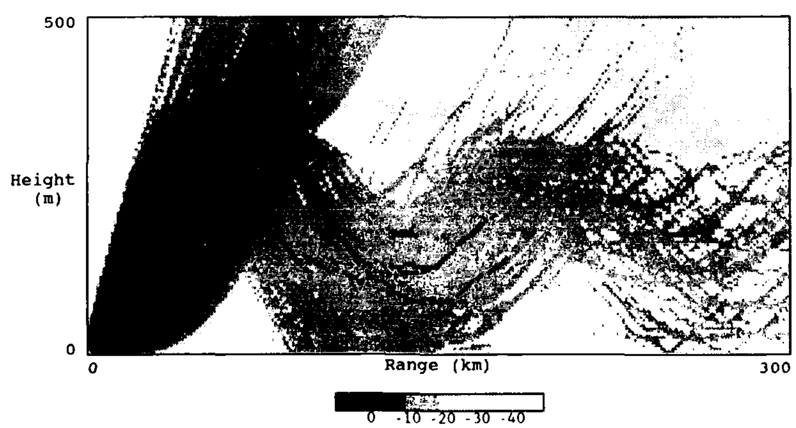


Figure 11: Field strength contours (dB) for the same system as in Figure 9. The ducting layer is corrugated (amplitude 25 m and wavelength 5 km).

DISCUSSION

L. FELSEN, US

Your PE code apparently does an adequate job (compared to measured data) for your particular applications, over a certain parameter range, but do you have any quantification of the errors? Over what ranges have you tested the algorithm against reference solutions (like normal modes) in range independent environments? To put it succinctly, how universal (apart from what you have tried so far) do you think the algorithm is?

In your reference to ray solutions, you referred only to the most primitive ray-trace forms, but there are ray codes that track phase and amplitude of the field, and also (via uniformization) penetrate ray trace shadow zones. Are you aware of these? (Of course, even those more refined versions have limitations, especially for rays that split at the upper duct boundary).

In one of your very striking color plots, there are oval-shaped regions that definitely suggest the presence of guided modes. Have you compared your code with well-trapping mode ducts with modal reference solutions, and over what distances do you feel that PE and MODE stay in step?

AUTHOR'S REPLY

We get good agreement between our PE code and one-dimensional mode theory methods at microwave frequencies, for ranges out to several hundred kilometers. Of course, mode theory methods can only be used with highly idealized refractivity profiles, so we have an experimental program of airborne measurements in order to validate the model in the real environment. We'll be happy if the predictions agree with the observations to within 6 dB! The color plot to which you refer has not been compared with mode results, but similar calculations have been.

We have taken a somewhat empirical approach to controlling numerical and model-generated errors, by simply using a fine enough integration grid to ensure convergence, and this appears to work for the problems that we have explored. Other PE workers have made more sophisticated analyses of the limits of the model, and more care would be required for wide angle problems or phase calculations at long range, for example.

We have in fact developed a "quantitative" ray model such as you have described which does a full two-dimensional integration of the ray equations. However, we could not find a way of automating the program so that it would cope adequately with "arbitrary" refractive index structures (the main problem being the splitting of ray families to which you referred); the program turned out to be slower than our PE code in any case for a full two-dimensional field strength diagram, so there seemed little point in pursuing this option.

ETUDE DE LA PROPAGATION DANS UNE ATMOSPHERE INHOMOGENE DANS LES DIRECTIONS HORIZONTALE ET VERTICALE PAR LA METHODE DE L'EQUATION PARABOLIQUE

Michel FOURNIER
Le Centre Thomson d'Applications Radars
6 rue Nieuport
78143 VELIZY-VILLACOUBLAY
FRANCE

SUMMARY

Operational programs calculating radar coverage are now available on ships having a meteorological capability.

A fundamental assumption of these programs is that only vertical variations of refraction index are known and consequently that the atmosphere is homogeneous along the propagation path in the horizontal direction.

Measurements made by the French Meteorological Office and by other countries established that this hypothesis may be inconsistent during significal laps of time.

Consequently it seems necessary to develop a tool to compare results of existing operational programs and those given by more sophisticated programs which are able to take into account variations of refractive index in vertical and horizontal directions.

In this paper, a resolution method of this problem, using the parabolic equation is described.

This tool has been used for exploitation of index of refraction map from in-situ measurements obtained during tests made over the Atlantic Ocean and the Mediterranean Sea. Exploitation of these data allows a study of the effect of horizontal index inhomogeneities of the atmosphere in horizonta direction.

1. INTRODUCTION

La propagation des ondes radioélectriques dans l'atmosphère est déterminée par l'indice de réfraction qui est sensible à la pression, à la température et à l'humidité de l'air. L'apparition de calculateurs embarquables a permis de mettre à la disposition des utilisateurs les moyens nécessaires pour effectuer une prédiction de la propagation en fonction de la connaissance des valeurs de l'indice de réfraction sur une zone donnée.

Actuellement, des programmes de calcul de la couverture radar sont opérationnels sur les bâtiments de surface [1-2] .

Ces programmes reposent sur la connaissance d'une seule coupe de l'indice atmosphérique en fonction de l'altitude et supposent que la loi de variation en altitude ainsi déterminée reste la même sur toute la zone considérée, ce qui revient à admettre que l'indice ne varie pas en fonction de la distance.

Des mesures faites à l'étranger et en France par la Météorologie Nationale montrent que cette hypothèse n'est pas toujours vérifiée. Dans certains cas, on peut constater des variations notables de l'indice atmosphérique en fonction de la distance [3-4] . Dans ces conditions, il a semblé opportun de mettre au point une méthode de calcul dans laquelle on peut tenir compte à la fois des variations de l'indice en fonction de l'altitude et de la distance. Cette méthode a été testée à l'aide de données fournies par la Météorologie Nationale et correspondant à plusieures campagnes de mesures effectuées en Méditerranée et en Atlantique.

2. CHOIX D'UNE METHODE DE CALCUL

Dans le cas où seule intervient une variation en fonction de l'altitude de l'indice atmosphérique, le problème est à symétrie sphérique et cette particularité peut être exploitée pour développer un modèle analytique. Dans le cas où l'indice de réfraction dépend à la fois de l'altitude et de la distance, cette symétrie sphérique n'existe plus et une méthode de calcul purement analytique ne peut être envisagée.

Une étude bibliographique a montré que seules trois méthodes peuvent être utilisées dans ce genre de problème :

- la méthode des modes couplés [5-8]
- la méthode des rayons [9-11]
- la méthode de l'équation parabolique [13-16] .

Les deux premières méthodes ont été exclues : la méthode des modes couplés à cause de sa complexité de mise en oeuvre et la méthode des rayons à cause de la difficulté intrinsèque qu'elle présente pour fournir une solution quantitative dans le domaine envisagé.

En revanche la méthode de l'équation parabolique est simple à mettre en oeuvre et sa parfaite adéquation au problème posé l'a fait retenir.

3. DESCRIPTION DE LA METHODE DE CALCUL

3.1 Théorie de l'équation parabolique

Cette méthode consiste à faire une approximation de l'équation d'onde en supposant que le champ se propage autour d'une direction privilégiée. Dans ces conditions, on ramène l'équation de propagation à une équation aux dérivées partielles du type parabolique qui permet d'analyser le champ autour de cette direction privilégiée. Ceci revient à l'approximation paraxiale de l'Optique. D'un point de vue physique, on néglige les ondes rétrodiffusées pour ne considérer que celles qui se propagent autour de la direction axiale. Une telle approximation est utile lorsque l'inhomogénéité du milieu peut être considérée comme faible dans le sens de la propagation ce qui est le cas pour l'atmosphère.

Les cas de la polarisation verticale et de la polarisation horizontale doivent être envisagés séparément. Dans le cas de la polarisation verticale on ne restreint pas la généralité du problème en admettant que la source est constituée par un dipôle électrique vertical, ce qui permet d'envisager un problème à symétrie cylindrique.

Dans ce cas, il est avantageux de s'intéresser au champ magnétique $\vec{H}$. Les équations de Maxwell permettent d'écrire :

$$\vec{\nabla} \times \vec{\nabla} \times \vec{H} - k^2 \vec{H} = \frac{\vec{\nabla}\varepsilon}{\varepsilon} \times \vec{\nabla} \times \vec{H} \qquad (1)$$
$$k^2 = \omega^2 \mu \varepsilon$$

expression où : ω est la pulsation de l'onde
μ est la perméabilité magnétique de l'atmosphère (égale pratiquement à celle du vide)
ε est la constante diélectrique de l'atmosphère. Celle-ci est une fonction de l'altitude et de la distance.

Compte tenu de la symétrie de révolution du champ électrique, $\vec{H}$ n'a qu'une composante horizontale et le problème peut se ramener à un problème scalaire plan.

L'équation (1) peut être particularisée dans un système de coordonnées sphériques dont l'origine est le centre de la terre et dont l'axe $\vec{Oz}$ est dirigé dans le sens du dipôle électrique vertical modélisant la source.

Le changement de variable :

$$H = \frac{u(r,\theta)}{r} \frac{1}{\sqrt{\sin\theta}} e^{-j k_0 a \theta}$$

a = rayon de la terre $\quad k_0 = k(a,0)$ = nombre d'onde sur la surface de la mer (relatif aux constantes électriques de l'atmosphère)

permet de s'affranchir des variations de grande amplitude du champ au voisinage de la source ainsi que des oscillations rapides en fonction de la distance.

Le terme $u(r,\theta)$ représente un terme d'atténuation dont les variations sont relativement lentes comparativement à la longueur d'onde. Après cette substitution (1) se ramène à une équation aux dérivées partielles de type elliptique pour le terme d'atténuation $u(r,\theta)$. Cette équation exacte peut être simplifiée - FOCK [12] a admis que le terme $\frac{\partial^2 U}{\partial \theta^2}$ est négligeable devant le terme $k_0 a \frac{\partial U}{\partial \theta}$. Cette approximation permet de transformer l'équation du type elliptique en une équation du type parabolique beaucoup plus facile à résoudre.

FOCK justifie cette approximation en comparant la solution analytique de l'équation parabolique dans le cas d'une atmosphère à indice constant avec la solution exacte du champ diffracté par la sphère terrestre obtenue par sommation d'une série modale.

Cette équation simplifiée s'écrit :

$$\frac{\partial^2 U}{\partial r^2} - 2j \frac{k_0}{a} \frac{\partial U}{\partial \theta} + k_0^2 \left(\frac{\varepsilon - \varepsilon_0}{\varepsilon_0} + 2 \frac{r - a}{a} \right) U = 0$$

On peut remarquer que cette équation est du second ordre dans le sens vertical et du premier ordre dans le sens horizontal.

Les variations de l'indice atmosphérique sont introduites par le terme $\underline{\varepsilon - \varepsilon_0}$ où ε est une fonction de l'altitude (r) et de la distance (θ). Le terme $\frac{2(r-a)\,\varepsilon_0}{a}$ revient à introduire une variation linéaire de l'indice de l'atmosphère qui tient compte d'une manière automatique de la rotondité de la terre tant que les altitudes considérées restent faibles par rapport à son rayon, ce qui est toujours le cas pour le type de problème considéré ici.

Les conditions de validité de l'approximation parabolique sont données par les relations ci-dessous :

$$\frac{1}{\varepsilon}\,\frac{\partial \varepsilon}{\partial \theta} \ll 2m \qquad \frac{1}{\varepsilon}\,\frac{\partial \varepsilon}{\partial h} \ll \frac{k_0}{m} \qquad m = \sqrt[3]{\frac{k_0 a}{2}}$$

$$\theta \gg \frac{1}{2m^2} \qquad k_0 a\,\frac{\partial U}{\partial \theta} \gg \frac{\partial^2 U}{\partial \theta^2}$$

Les deux premières relations indiquent respectivement une contrainte sur le gradient de l'indice en fonction de la distance et en fonction de l'altitude.

La troisième relation traduit une contrainte sur la distance par rapport à la source.

Des études postérieures aux travaux de FOCK ont montré que la quatrième relation est toujours vérifiée si les rayons ne sont pas trop obliques par rapport à la direction horizontale (leur inclinaison ne doit pas dépasser 20°) [16] . Cette contrainte entraîne que le terme d'atténuation n'est pas calculé dans la zone zénithale. Ceci n'est pas très important dans le cadre de la présente étude car on s'intéresse surtout aux grandes distances. La couverture dans la zone zénithale peut être complétée par une méthode de rayons qui est bien adaptée aux courtes distances concernées. Dans les cas rencontrés en pratique, les gradients d'indice ne sont que de quelques unités N par mètre (les unités N sont données par $N = (n-1)\,10^6$, expression où n est l'indice atmosphérique) et les bornes données ci-dessus ne constituent pas une contrainte réelle. La contrainte sur la distance ne joue pas non plus, si bien que l'approximation parabolique se trouve amplement justifiée. Pour résoudre le problème différentiel qui se pose, il convient de disposer d'une condition initiale (problème de CAUCHY) et de deux conditions aux limites sur les bornes inférieure et supérieure du domaine de propagation.

La mer ayant une conductivité élevée, son influence peut être caractérisée par une impédance de surface, ce qui constitue l'approximation de LEONTOVITCH [17] . Elle consiste à écrire que le champ électrique tangentiel induit un courant proportionnel sur la surface :

$$\vec{E_t} = \sqrt{\frac{\mu_m}{\varepsilon_m}}\;\vec{H_t} \times \vec{n}$$

expression où $\vec{n}$ est la normale à la surface dirigée vers le centre de la terre.

μ_m et ε_m sont les constantes électriques de l'eau de mer. ε_m est ici la constante diélectrique complexe.

En terme d'atténuation, la relation précédente s'écrit :

$$\frac{\partial U}{\partial r} - \frac{j\,k_0}{\sqrt{\eta}}\,U = 0$$

expression où : $$\eta = \frac{\varepsilon}{\varepsilon_0}\left(1 - j\,\frac{\sigma}{\omega\varepsilon}\right)$$

pour l'eau de mer $\frac{\varepsilon}{\varepsilon_0} = 80$ et $\sigma \simeq 4.1$ mho/m

L'approximation de LEONTOVITCH apporte une grande simplification au problème, car seul est pris en compte le calcul du champ au-dessus de la surface de la mer. En toute rigueur, il faudrait aussi traiter le cas des ondes qui pénètrent le dioptre mer-atmosphère, mais aux fréquences considérées ces ondes sont très rapidement atténuées, ce qui justifie l'approximation.

La condition aux limites sur la frontière supérieure du domaine s'écrit :

$$\lim_{r \to \infty}\left(\frac{\partial U}{\partial r} + j\,k_0\,U\right) = 0$$

Elle traduit simplement le fait que l'onde doit s'éloigner de la source lorsque le point d'observation est situé loin de celle-ci.

Dans le cas de la polarisation horizontale on admet que la source est constituée par un dipôle magnétique vertical. Dans ce cas il convient de s'intéresser au champ électrique $\vec{E}$ qui ne possède qu'une composante horizontale. Par un changement de variable identique à celui qui a été effectué dans le cas de la polarisation verticale et avec les mêmes hypothèses simplificatrices, on obtient exactement la même équation aux dérivées partielles relativement au terme d'atténuation. Seule change la condition aux limites sur la surface qui s'écrit maintenant :

$$\frac{\partial U}{\partial n} - j k_o \sqrt{\eta}\, U = 0$$

La condition aux limites sur la frontière supérieure du domaine est la même que dans le cas de la polarisation verticale.

3.2 Résolution numérique de l'équation parabolique

Un grand nombre de méthodes numériques permettent la résolution d'équations aux dérivées partielles de type parabolique. La plupart des méthodes proposées utilisent une représentation aux différences finies de l'équation originale [18-23] . Une autre approche qui jouit d'une grande popularité fait appel à la Transformée de Fourier-Rapide [24] . Après avoir comparé les temps de calcul de ces deux approches possibles et afin de bien maîtriser les conditions aux limites sur les frontières du domaine, les calculs numériques ont été effectués à l'aide d'une représentation aux différences finies de l'équation parabolique et plus précisément par la méthode de CRANK-NICHOLSON.

Le schéma aux différences obtenu se ramène à un système linéaire du type :

$$A V^{m+1} = B V^{m}$$

où V^m représente le vecteur correspondant à une verticale de la grille de maillage au pas m en distance. Les deux matrices A et B sont tridiagonales.

$$V^{m+1} = L_m V^m \qquad \text{où} \qquad L_m = A^{-1} B$$

On passe de la solution au pas m en distance à la solution au pas $m + 1$ par application d'un opérateur linéaire sur le vecteur solution au pas correspondant à l'ensemble des valeurs de la fonction sur une verticale de la grille de maillage. L_m est obtenu par la méthode du "double balayage" qui résoud le système linéaire en mettant à profit la structure tridiagonale des matrices A et B [22] . Cette méthode de calcul est rapide, précise et facile à mettre en oeuvre sur calculateur.

On peut remarquer que :

$$V^m = L_m L_{m-1} \cdots\cdots L_1 V^0$$

la solution au rang m est obtenue en itérant m fois l'opérateur de passage, la précision de la solution est donc conditionnée par la précision de la solution initiale et par la précision sur l'opérateur de passage ($L_m L_{m-1} \cdots\cdots L_1$) qui permet d'obtenir le vecteur V^m à partir du vecteur V^0 . En particulier, le vecteur initial qui est obtenu par une méthode autre que celle de l'équation parabolique (méthode des modes, des rayons ou méthode WKB), doit être fourni avec une très bonne précision (en particulier sur la phase), car il conditionne toute les évaluations suivantes.

3.3 Mise en oeuvre de l'algorithme

Pour mettre en oeuvre l'algorithme de résolution qui vient d'être décrit, il est nécessaire :

- de disposer d'une solution initiale,
- de mailler convenablement le domaine de calcul,
- de limiter le domaine vers le haut.

Dans les essais qui ont été effectués, la solution initiale a été obtenue par une méthode de rayons généralisée en admettant que le processus de calcul démarre suffisamment près de la source pour pouvoir faire une approximation locale de terre plate. Cette solution initiale tient compte de l'ouverture en site du faisceau correspondant à la source d'émission. Pour mailler convenablement le domaine de calcul, il est nécessaire de tenir compte du caractère rapidement oscillant de la solution au fur et à mesure que l'on s'éloigne de l'axe du faisceau. Des considérations simples basées sur la solution donnée par l'Optique Géométrique dans le cas d'une atmosphère homogène ont permis d'établir la règle de maillage suivante :

$$\Delta s \simeq \frac{\lambda}{5\alpha^2} \qquad \text{pour le pas en distance}$$

$$\Delta h \simeq \frac{\lambda}{10\alpha} \qquad \text{pour le pas en altitude}$$

dans ces formules, α est la demi-ouverture du faisceau.

Pour limiter le domaine de calcul vers le haut il convient d'écrire une condition aux limites appropriée sur la frontière supérieure de la grille de maillage. L'idée la plus naturelle consiste à écrire qu'au sommet du domaine de calcul l'impédance est égale à celle de l'atmosphère à l'altitude considérée ce qui revient à traduire sur le plan numérique la condition de SOMMERFELD qui est introduite lors de la mise en équation du problème.

Les essais basés sur cette approche ont montré que dès que le faisceau atteint la frontière supérieure il y a une réflexion qui engendre un terme perturbateur se propageant vers les altitudes décroissantes et qui vient rapidement altérer la solution lorsqu'on s'éloigne de la source.

En vue d'amortir ce terme parasite, une méthode heuristique a été mise en oeuvre, elle consiste à admettre que le fait d'annuler un certain nombre de termes de la solution vers le haut n'altère pas celle-ci pour les pas de calcul suivants en distance. Tout se passe alors comme si l'on restait en espace libre en rétrécissant artificiellement le faisceau vers le haut au fur et à mesure que l'on progresse en distance.

La méthode est mise en oeuvre de la manière suivante :

- le domaine de calcul est agrandi vers le haut par adjonction d'une zone tampon
- la condition aux limites évoquée ci-dessus est appliquée à la frontière supérieure du domaine étendu
- à chaque étape en distance, la solution est calculée puis multipliée par une fonction telle que les termes se trouvant dans la zone tampon soient écrasés et que soient conservés les termes correspondant au domaine utile.

Dans un premier temps, une fonction rectangle (égale à 1 dans la zone utile et égale à 0 dans la zone tampon) a été utilisée. L'expérience a montré que la solution obtenue était légèrement perturbée. Cette perturbation disparaît en introduisant une zone de transition faisant passer la valeur de la fonction progressivement de 1 à 0 dans le voisinage de la frontière séparant le domaine utile de la zone tampon. Après plusieurs essais, il est apparu qu'un bon compromis consiste à agrandir le domaine de calcul dans un rapport 1,5 en hauteur et à adopter une fonction d'apodisation du type :

$$F(x) = \frac{1}{2}\left[1 + Erf\left(\frac{1-x}{\sigma\sqrt{2}}\right)\right]$$

le coefficient σ étant choisi de façon à ce que cette fonction soit pratiquement nulle pour $x \geqslant 1{,}35$.

Le temps de calcul croît linéairement en fonction de la distance avec cependant l'inconvénient de multiplier le nombre d'opérations arithmétiques dans un rapport 1,5 par rapport à ce qu'il serait strictement nécessaire si l'on pouvait se limiter au domaine de calcul utile.

3.4 Validation de la méthode de calcul

Une solution analytique rigoureuse de l'équation parabolique peut être trouvée dans le cas d'une atmosphère homogène et dans le cas d'une atmosphère ayant un indice variant linéairement en fonction de l'altitude. Dans le cas où le point d'observation est situé en dessous de l'horizon géométrique, la solution analytique est fournie par une série modale dont le terme général est un produit de fonctions d'AIRY. Dans le cas où le point d'observation est situé au-dessus de l'horizon, la solution est obtenue par un développement asymptotique et coïncide avec la solution donnée par l'Optique Géométrique. La solution dans tout le domaine est obtenue en raccordant judicieusement les deux solutions dans la zone de transition correspondant à la limite de validité de chacune des méthodes de calcul. Ces solutions analytiques ont permis de valider la méthode de calcul exposée en comparant le calcul analytique à la solution obtenue en résolvant numériquement l'équation parabolique.

La figure 1 donne un exemple de comparaison entre la solution numérique fournie par la méthode de l'équation parabolique et la solution analytique. On peut constater qu'il existe une très bonne concordance dans l'ensemble entre ces deux solutions. La figure 2 montre les possibilités de la méthode en envisageant un conduit d'évaporation de 20 mètres d'épaisseur à profil bilinéaire (dN_1/dh = -0.3 N/mètre ; dN_2/dh = -0.039 N/mètre) à une fréquence de 10 GHz et pour lequel on fait varier l'altitude de la source.

4. EXPLOITATION POUR QUELQUES CAS TYPES

Afin d'évaluer le degré d'hétérogénéité de l'indice atmosphérique, la Météorologie Nationale a effectué un certain nombre de campagnes de mesures au-dessus de l'Atlantique et de la Méditerranée. Ces mesures ont conduit à l'élaboration de cartes donnant l'indice atmosphérique en fonction de l'altitude et de la distance. Les résultats sont stockés sur des bandes magnétiques, et chaque carte couvre une zone de 130 km en distance avec un pas de 1 km et 1500 m en altitude avec un pas de 25 m. L'algorithme de calcul de la couverture radar décrit ci-dessus a été évalué sur un certain nombre de configurations types extraites de ces fichiers magnétiques et choisies en accord avec

la Météorologie Nationale. Le pas desmesures effectuées in-situ est trop grand pour une exploitation directe dans les calculs de résolution de l'équation parabolique et il a été nécessaire d'effectuer une interpolation linéaire à deux dimensions pour obtenir, à partir de la grille météorologique, une grille de valeurs de l'indice correspondant à la grille descalculs numériques. Dans chaque cas il est effectué une comparaison entre l'atmosphère translatée (la première coupe verticale de la carte d'indice est reproduite à tous les pas suivants en distance, ce qui conduit à une atmosphère homogène dans le sens de la propagation) et l'atmosphère réelle obtenue par les mesures effectuées in-situ.

4.1 Vol CARMEN-14 (figure 3)

Le vol CAMEN 14 se caractérise par un conduit très net à basse altitude (entre 100 et 300 m). Cette disposition a été mise à profit pour étudier le cas où la source est nettement au-dessus du conduit. Les figures relatives respectivement à l'atmosphère réelle et à l'atmosphère translatée en distance présentent les points suivants :

- dans les deux cas, il y a tendance à la création d'un conduit faible à moyenne altitude (entre 500 m et 1000 m)
- une partie de l'énergie du faisceau qui traverse le conduit est réfléchie sur la surface de la mer puis se trouve concentrée entre la surface de la mer et le plancher du conduit
- cette tendance est nettement plus marquée dans le cas de l'atmosphère translatée.

4.2 Vol CARMEN-19 (figure 4)

Le vol CARMEN 19 présente une anomalie dans les courbes d'iso-indice caractérisée par un effondrement sur une tranche distance se situant entre 50 et 90 km par rapport à l'origine. Le but des exploitations était de vérifier si cette anomalie a une influence notable sur la propagation. Les figures montrent qu'il n'y a pas de différence sensible du point de vue qualitatif entre l'atmosphère translatée en distance (donc sans anomalie d'indice) et l'atmosphère réelle. Le phénomène a donc peu d'influence sur la propagation. Un complément d'essais a été effectué sur une source à basse altitude et conduit à la même conclusion. Ceci provient vraisemblablement du fait que l'anomalie est située trop loin de la source pour apporter une modification sensible aux conditions de propagation (compte tenu de la courbure terrestre l'énergie contenue dans le faisceau passe au-dessus de la zone d'effondrement et sa trajectoire est peu affectée).

4.3 Vol CARMEN-21 (figure 5)

Ce vol a été choisi comme extrêmement représentatif d'hétérogénéités de l'atmosphère en altitude et en distance. En effet, l'examen de la carte d'indice montre qu'il existe un fort gradient à une altitude variant entre 700 m et 1100 m traduisant un conduit de propagation hétérogène. Ce vol a été systématiquement exploité pour diverses altitudes de la source et diverses fréquences. Les résultats présentés ici sont relatifs à une source située à 500 mètres et émettant à une fréquence de 3 GHz. L'examen de la figure 5 montre les grandes différences qui peuvent exister entre la propagation en atmosphère réelle et la propagation en atmosphère translatée. En se basant sur une seule coupe translatée en distance, le calcul prédit un conduit de propagation extrêmement fort alors que la situation réelle est beaucoup plus complexe, mettant en évidence un conduit plus faible à 500 mètres ainsi qu'une surpropagation vers les altitudes élevées qu'il était impossible à prédire avec la connaissance d'une seule coupe verticale de la carte de l'indice atmosphérique.

5. CONCLUSION

La présente publication décrit une méthode de calcul de la couverture radar adaptée au cas où l'indice de réfraction de l'atmosphère dépend à la fois de l'altitude et de la distance. Cette méthode est basée sur la résolution numérique d'une équation aux dérivées partielles du type parabolique qui constitue une approximation raisonnable de l'équation de propagation. L'algorithme choisi fournit un outil d'investigation pour l'étude de la couverture radar dans le cas d'une configuration quelconque de l'indice atmosphérique. Il s'agit d'un moyen d'étude et non pas d'un programme opérationnel, car les temps de calcul sont trop longs pour envisager son emploi dans un contexte opérationnel. Cependant cet algorithme est parfaitement adapté à une implémentation sur un processeur vectoriel, ce qui réduirait le temps de calcul dans un rapport d'environ vingt et permettant alors un emploi à des fins opérationnelles.

La mise en oeuvre de cet algorithme sur des mesures effectuées in-situ permet de mettre en évidence, dans certains cas, une différence importante entre le champ calculé à partir d'une seule coupe verticale de l'indice et celui calculé pour une atmosphère réelle. Il conviendrait donc de pousser plus à fond les investigations afin d'obtenir des données statistiques permettant de déterminer dans quelle mesure un seul relevé météorologique permet sur le plan opérationnel de rendre compte d'une manière fiable de la réalité. Ceci pose évidemment, d'une manière implicite, le problème de l'obtention dans un contexte opérationnel de la carte de l'indice atmosphérique sur une zone étendue.

REMERCIEMENTS

Cette étude a été réalisée sous contrat de la Direction des Recherches, Etudes et Techniques (Délégation Générale pour l'Armement).

REFERENCES

[1] H.V. HITNEY - J.H. RICHTER,
Integrated Refractive Effects Prediction System (IREPS) URSI Commission F Proceedings. La Baule (France) 28 avril - 6 mai 1977 - pp. 55-60

[2] G.B. BAUMGARTNER, H.V. HITNEY, R.A. PAPPERT,
Duct propagation modelling for the integrated - refractive - effects prediction system (IREPS). IEE Proceedings, Vol. 130, Part. F, n° 7, december 1983 - pp. 630-642

[3] B.R. BEAN - E.J. DUTTON,
Radio Meteorology US Department of Commerce National Bureau of Standards Monograph 92. Issued March 1, 1966

[4] B. STRAUSS,
Evaluation de la hauteur du conduit troposphérique d'évaporation. AGARD Conference Proceedings n° 346 - 4-7 october 1983. Characteristics of the lower atmosphere influencing radio-wave propagation pp. 20-1 20-8. AGARD 7, rue Ancelle 92200 Neuilly-sur-Seine - FRANCE

[5] S.H. CHO - J.R. WAIT,
Analysis of microwave ducting in unhomogeneous troposphere. Pure and Applied Geophysics 116 - 1979 - pp. 1118 - 1142

[6] F.S. CHIEROTH, A. NAGL, H. UBERALL, R.D. GAVES, G.L. ZARUR,
Mode coupling in a sound channel with range dependent parabolic velocity profile. J. Acoust. Soc. Am. 64 (4) - Oct. 1978 pp. 1105-1112

[7] C. ALLAN BOYLES,
Coupled mode solution for a cylindrically symmetric oceanic waveguide with a range and depth dependant refractive index and a time varying rough sea surface. J. Acoust. Soc. Am. 73 (3) March 1983 - pp. 800-805

[8] C.A. BOYLES - L.B. DOZIER,
Coupled mode theory for an unhomogeneous oceanic wave guide with a time varying, randomly rough sea surface. 11ème Congrès International d'Acoustique. PARIS 19ème - 27 juillet 1983 - Revue d'Acoustique Hors Série - Volume 2 - pp. 435-439

[9] M. BORN - E. WOLF,
Principles of Optics. 1959, Pergamon Press London-New York-Paris-Los Angeles

[10] D.E. KERR,
Propagation of short Radio-waves - pp. 41-58 Dover publications Inc. New York 1965

[11] I.P. SHKAROFSKY - S.B. NICKERSON,
Computer modeling of multipath propagation : Review of ray-tracing techniques Radio Science volume 17, number 5, pp. 1133-1158, September - October 1982

[12] V.A. FOCK,
Electromagnetic Diffraction and Propagation Problems 1965 Pergamon Press Oxford - London - Edinburgh - New York - Paris - Frankfurt.

[13] H.W. HO, H.P. SKURA, J.W. SARI,
Surface Duct radar propagation modeling 1983 IEEE - pp. 177-183.

[14] H.W. KO - J.W. SARI - M.E. THOMAS - P.J. HERCHENROEDER and P.J. MARTONE
Anomalous propagation and radar coverage through unhomogenous atmospheres AGARD Conference Proceedings n° 346 Characteristics of the lower atmosphere influencing radio-wave propagation 7, rue Ancelle 92200 Neuilly-sur-Seine FRANCE pp. 25-1 - 25-13 4-7 october 1983

[15] H.W. KO - et al,
Anomalous microwave propagation through atmospheric ducts Johns Hopkins APL Technical Digest Volume 4, Number 1, 1983 - pp. 12-26

[16] SUZANNE T - Mc DANIEL,
Propagation of normal mode in the parabolic approximation J. Acoust. Soc. Am., Volume 57, number 2, February 1975 pp. 307-31.

[17] L. LANDAU et E. LIFCHITZ,
Electrodynamique des milieux continus, Edition MIR MOSCOU 1969 pp. 365-371

[18] N. GASTINEL,
Mathématiques pour l'informatique 2 - Equations différentielles Librairie Armand Colin - PARIS 1970

[19] J. LEGRAS,
Méthodes et techniques de l'analyse numérique DUNOD Paris 1971

[20] G.E. FORSYTHE - W.R. WASOW,
Finite difference methods for partial differential equations. 1960 John Wiley and Sons Inc. New-York - London - Sydney

[21] I.S. BEREZIN - N.P. ZHIDKOV,
Computing methods - Volume 2 1965. Pergamon Press Oxford - London - Edinburg New-York - Paris - Frankfurt

[22] S. GODOUNOV et V. RIABENKI,
Schémas aux différences Editions MIR MOSCOU 1973 - Chapitre 3, pp. 55-67

[23] M. CANDEL,
Acoustique Ecole Centrale des Arts et Manufactures Option Océan et Air Espace - Volume 2, pp. 65-119.

[24] R.H. HARDIN - F.D. TAPPERT,
Application of the split-step Fourier Method to the numerical solution of non linear and variable coefficient wave equation. SIAM rev. 115 - 1973 - pages 423. Proceeding of the Eighth International Congress on Acoustics (Goldcrest, London 1974), Volume 2, page 452.

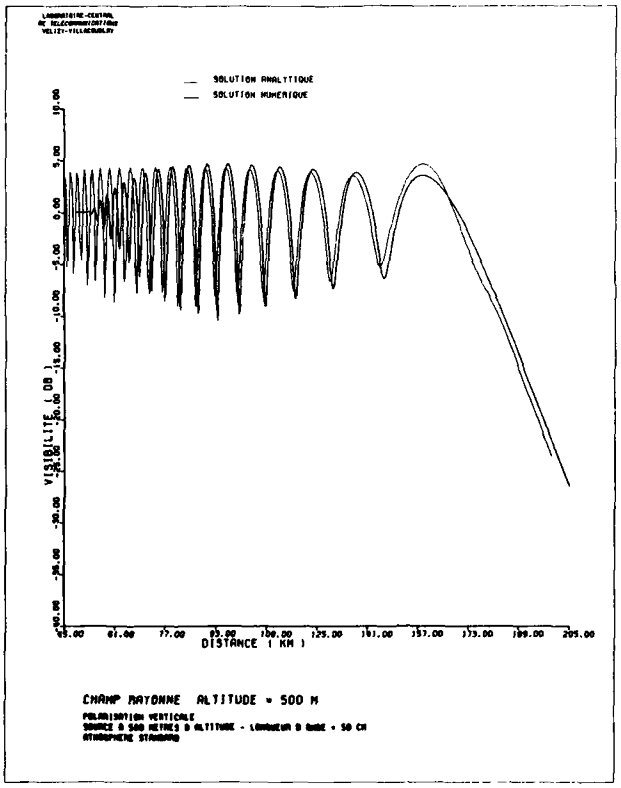


Figure n° 1 - Comparaison de la solution numérique avec la solution analytique

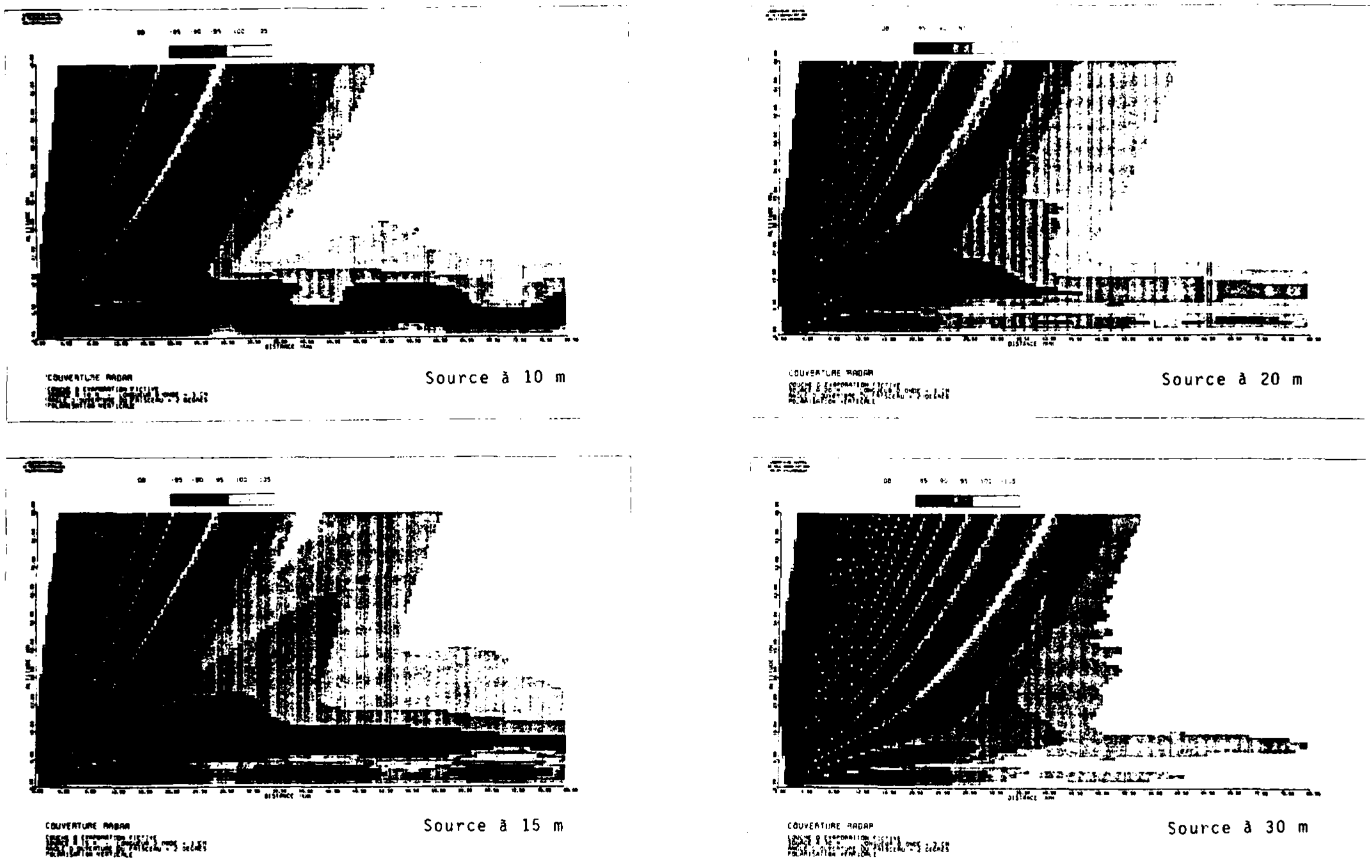
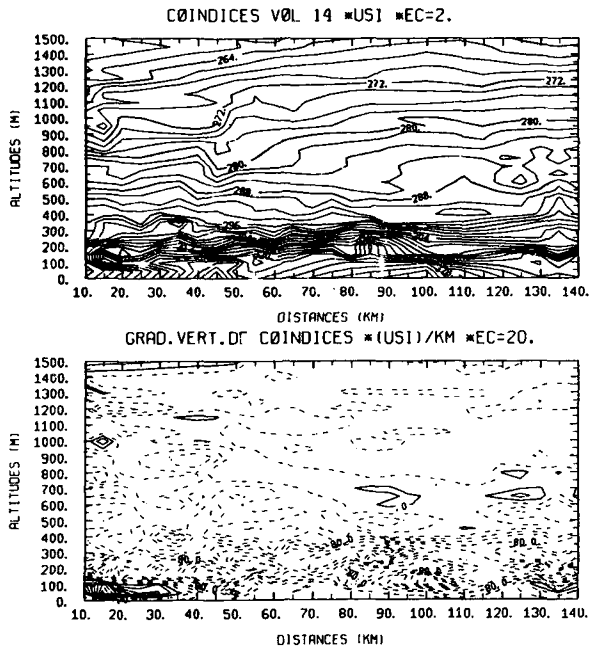


Figure n° 2 - Etude d'un conduit bilinéaire (h_c = 20 m λ = 3 cm)



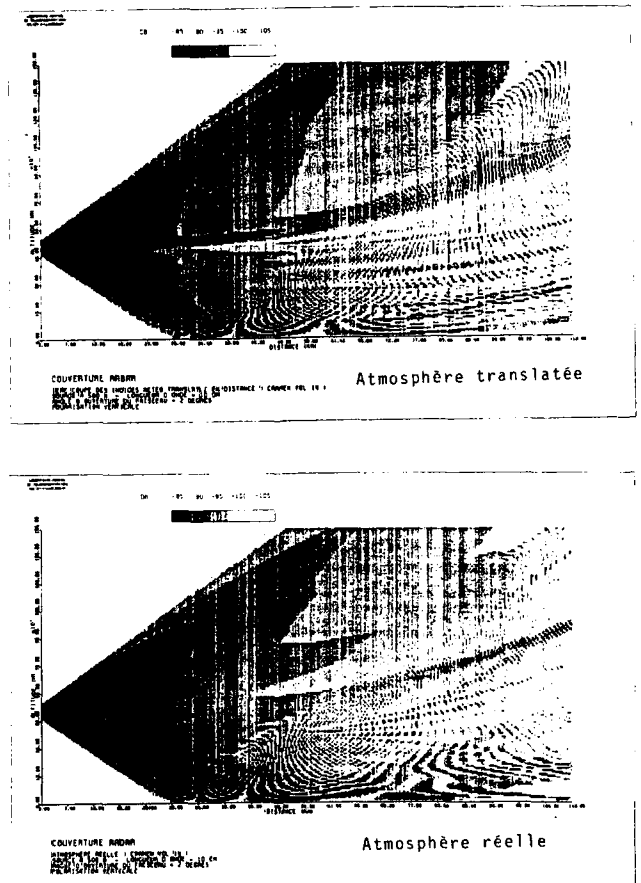
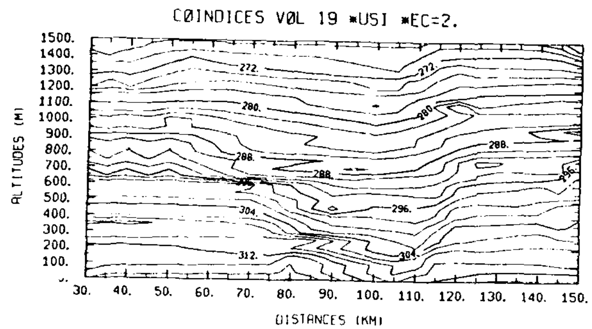


Figure n° 3 - Vol CARMEN 14

Source à 500 m λ = 10 cm



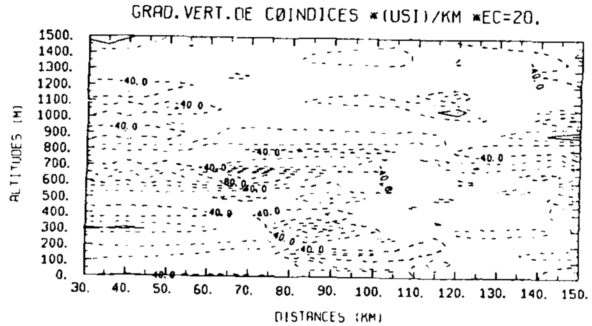
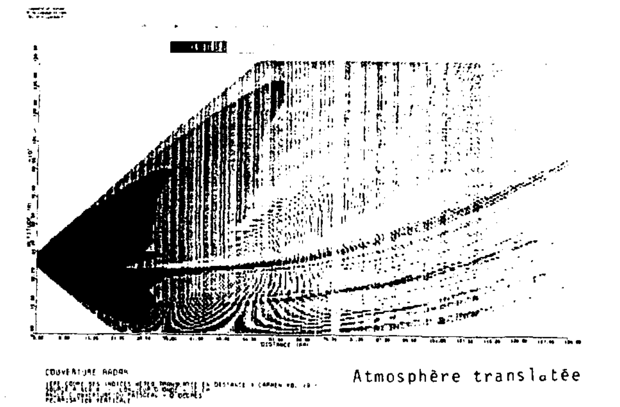
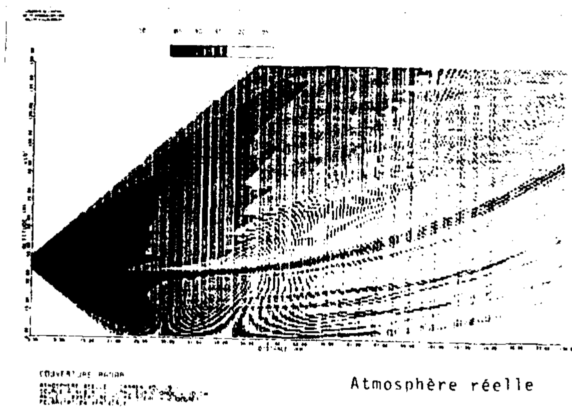


Figure n° 4 - Vol CARMEN 19

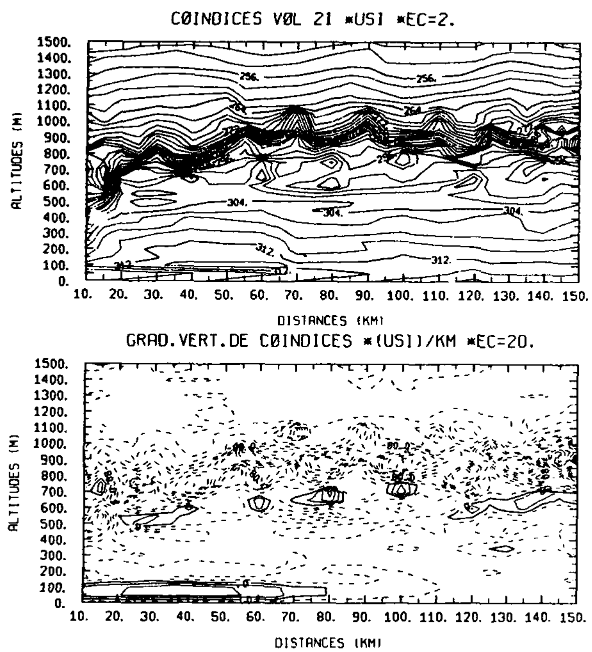
Source à 400 m λ = 10 cm



Atmosphère translatée



Atmosphère réelle



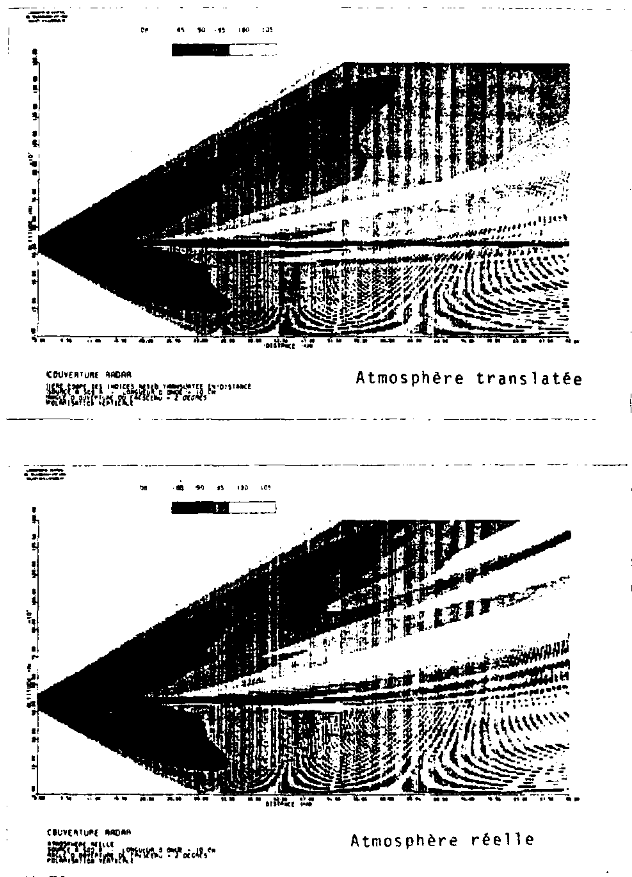


Figure n° 5 - Vol CARMEN 21

Source à 500 m λ = 10 cm

DISCUSSION

L. FELSEN, US

In attempting to satisfy the radiation condition on your finite domain, you apparently define some sort of transitional vector. Have you considered the possibility of using the "on-surface radiation condition," which has been introduced in diffraction problems and which simulates the radiation condition on a closed surface that is not in the far zone? Do you think it might work for your problem?

AUTHOR'S REPLY

In a first step to obtain a finite domain of calculus, we used the radiation condition at the boundary top of this domain. The results were catastrophic: a reflecting wave coming down from this boundary perturbed the solution. To avoid this strong perturbation, we tried to use an absorbing boundary condition. For this we referred to B. Engquist and A. Majda in "Absorbing Boundary Conditions for the Numerical Simulation of Waves" (Mathematics of Computation, Volume 31, Number 139, July 1977, Pages 629-651). We did not obtain significant results in this way: the reflected wave remained strong probably because of the very small angle between the incident wave and the boundary top of the domain. Consequently, we did not continue in this way and the heuristic method which is described in the paper was used and worked well. It seems probable that all possibilities of the paper of Engquist and Majda have not yet been used. In our work, the possibility of using the "on surface radiation condition" was not considered. It may be interesting to explore the way to obtain a more rigorous solution to the problem of limiting the area of computation. This may be a promising way for complementary work.

PROPAGATION MODELING FOR SOME HORIZONTALLY VARYING TROPOSPHERIC DUCTS

Richard A. Pappert
Naval Ocean Systems Center
Ocean and Atmospheric Sciences Division
San Diego, CA 92152-5000

SUMMARY

For propagation assessment in a maritime environment, the assumption of horizontal homogeneity (i.e., invariance of the refractivity structure along the path of propagation) is generally adequate [Hitney et al., 1985]. Nevertheless, theoretical methods are available to handle special cases when the assumption fails. For example, horizontal inhomogeneity can in principle be treated by mode conversion methods or by parabolic equation solvers. In this paper theory for a slab mode conversion model is discussed within the context of modified refractive index formalism and applied to several hypothetical laterally inhomogeneous problems including elevated layers as well as shallow evaporative layers. Results of the mode conversion calculations are compared with parabolic equation results generated concurrently by Ko and Burkom [1987]. Most of the comparisons between mode conversion and parabolic equation results are in reasonable agreement. These comparisons, along with the ease of implementation and speed of the parabolic equation calculations relative to the mode conversion analysis, strongly supports the superiority of the parabolic equation method for systems studies involving laterally inhomogeneous tropospheric layering.

INTRODUCTION

It is well known that the presence of layered tropospheric refractivity structure significantly influences radio wave propagation of frequencies greater than about 30 MHz to the extent that beyond the horizon ducted fields may be many tens of dB above the troposcatter fields [e.g. Kerr, 1951; Pappert and Goodhart, 1977]. Waveguide formalism [e.g. Budden, 1961; Wait, 1970] has served as the basis for many numerical studies and they have generally been predicated upon the assumption of a laterally homogeneous guide. That is, a guide for which the vertical refractivity is invariant along the path of propagation. As discussed by Hitney et al. [1985], for propagation prediction purposes in a maritime environment this assumption appears to be adequate most of the time. Nevertheless, it is legitimate to inquire about the theoretical significance of lateral inhomogeneities and several tools are available to handle such situations. When lateral variation is sufficiently slow, waveguide modes do not interact significantly and each mode can be tracked separately [Midler, 1969]. This is the adiabatic or WKB approximation. More generally, lateral inhomogeneity is treated by mode conversion techniques [Cho and Wait, 1978; Wait, 1980] or by parabolic equation solvers [Tappert, 1977, Ko et al., 1983; Dockery and Konstanzer, 1987; Ko et al., 1988]. Both mode conversion techniques and parabolic equation solvers have been developed to a high degree of sophistication by the underwater sound community. Despite this, the methods have been less frequently used in tropospheric ducting work. The principal reason for this is probably because the detailed information relating to lateral inhomogeneity of the duct which is necessary for meaningful tropospheric modeling is rarely available. Moreover, tropospheric ducts are deviations from the norm, whereas nonuniform ocean topology, for example, is the norm.

Although mode conversion methods become very unwieldly with increasing frequency as the system becomes more and more multimoded, they do offer a basis for comparison with alternative methods. Thus, the primary purpose of this study is to compare mode conversion results for some hypothetical laterally inhomogeneous tropospheric waveguide environments with the corresponding parabolic equation results concurrently generated by Ko and Burkom [1987]. The waveguide environments consist of two examples of elevated layers first considered by Cho and Wait [1978] and several horizontally inhomogeneous evaporation duct environments. The mode conversion analysis is based on a slab model of horizontal inhomogeneity and draws heavily upon an undocumented waveguide program, "MLAYER", which makes use of an ingenious root finder developed by Morfitt and Shellman [1976]. The latter permits the location of all modes with attenuation rates less than some preassigned value. "MLAYER" is a multi-layer extension of the trilinear program "XWVG" developed by Baumgartner [1983].

In the following section the slab mode conversion model is succinctly described. It is developed fully within the context of modified refractive index formalism. Since they play a crucial role in the mode conversion process, height gain integral evaluations are discussed in section III. Formulas for signal level relative to free space and for path loss are given in Section IV. Mode conversion and parabolic equation results for layers treated by Cho and Wait are presented in Section V and results for some evaporation duct environments are given in Section VI. Conclusions are summarized in Section VII.

THE SLAB MODEL AND MODE CONVERSION COEFFICIENTS

Throughout this study a right handed rectangular coordinate system (x, y, z) is use - z the plane of propagation and z the vertical coordinate in an earth flattened geometry. Positive z is directed into the troposphere and $z = 0$ corresponds to ground level. Invariance in y is assumed. As depicted in Figure 1, a slab model is used to represent inhomogeneity in the x direction. Within each slab there is no x dependence of the refractivity profile. Reflections associated with horizontal inhomogeneities are neglected. Some justification for this follows from the fact that reflections associated with a single junction can be shown to depend upon differences in propagation constants of the modes on each side of the junction and from the fact that in tropospheric applications those differences are very small.

By assuming an unit amplitude wave in mode m in the transmitter region (slab 1) and as just mentioned, by neglecting reflections from the horizontal or lateral inhomogeneities, the horizontal electric field, E_y (to good approximation the model is applicable to vertical

polarization with the magnetic field H_y replacing E_y) of the rf wave in slab 1 is

$$E_{ym}^{(1)} = \exp(-ik_o\beta_m^{(1)}x)\ f_m^{(1)}(z) \tag{1}$$

The superscript denotes slab number, m is a mode index, k_o is the free space wavenumber and β a modal eigenvalue. The height gain function $f_m^{(p)}(z)$ satisfies the reduced wave equation

$$\frac{d^2 f_m^{(p)}}{dz^2} + k_o^2(n_p^2(z) - (\beta_m^{(p)})^2)f_m^{(p)} = 0 \tag{2}$$

where n_p is the height dependent modified refractive index for the p^{th} slab. The square of n_p is approximated by linear segmentation so that the height gains are expressible in terms of Airy functions in each linear segment. The height gains are normalized such that

$$\int_\Omega (f_m^{(p)})^2 dz = 1 \tag{3}$$

where the contour Ω goes from $-\infty$ to ∞ in such a way that the integral converges. For leaky modes that necessitates integrating into the complex plane [Budden, 1961]. Moreover, it is readily established that the height gains for non degenerate modes are orthogonal. That is,

$$\int_\Omega f_m^{(p)}(z)\ f_k^{(p)}(z)\ dz = \delta_{mk} \tag{4}$$

where δ is the Kronecker delta.

In slabs $p \neq 1$ the y component of the electric field of the rf wave may be written as

$$E_{ym}^{(p)} = \exp(-ik_o\beta_m^{(1)}x_2) \sum_j^{J_p} a_{jm}^{(p)} \exp(-ik_o\beta_j^{(p)}(x-x_p))f_j^{(p)}(z) \tag{5}$$

where p is the slab index and J_p denotes the number of modes in slab p required to represent the field and the $a_{jm}^{(p)}$'s represent the cumulative conversion coefficients from mode m, incident on the transition in region 1, to mode j in the p^{th} slab. It is emphasized that the $a_{jm}^{(p)}$'s so defined, contain the effect of reconversion associated with all slab junctions for which $x_m \leq x_p$. Continuity of E_y at junction 2 yields

$$\sum_j^{J_2} a_{jm}^{(2)}\ f_j^{(2)}(z) = f_m^{(1)}(z) \tag{6}$$

Multiplying through by $f_r^{(2)}(z)$ and integrating over the contour Ω yields

$$a_{rm}^{(2)} = I_{r,m}^{2,1} \tag{7}$$

where

$$I_{r,m}^{p,q} = \int_\Omega f_r^{(p)}(z)\ f_m^{(q)}(z)dz \tag{8}$$

Repeating the continuity requirements for E_y at junctions $p > 2$ gives

$$a_{rm}^{(p)} = \sum_j^{J_{p-1}} a_{jm}^{(p-1)} \exp(-ik_o\beta_j^{(p-1)}(x_p-x_{p-1}))\ I_{r,j}^{p,p-1}; \quad p > 2 \tag{9}$$

Equations (7) through (9) permit the determination of the cumulative mode conversion coefficients.

HEIGHT GAIN INTEGRAL EVALUATIONS

A crucial step in the mode conversion program is the evaluation of height gain integrals given by Equation (8). When linear layers of n^2 in adjacent slabs are parallel, then the integral (8) can be evaluated analytically. Let a and b be two points (with a the lower point) which lie within the boundaries of parallel linear layers of n^2 in two adjacent slabs, r and s. The height gains in these slabs satisfy the equations (m and k are mode indices)

$$\frac{d^2 f_m^{(r)}}{dz^2} + k_o^2(n_r^2(z) - (\beta_m^{(r)})^2) f_m^{(r)} = 0 \qquad (10)$$

$$\frac{d^2 f_k^{(s)}}{dz^2} + k_o^2(n_s^2(z) - (\beta_k^{(s)})^2) f_k^{(s)} = 0 \qquad (11)$$

If the reduced wave Equation (10) is multiplied by $f_k^{(s)}(z)$ and Equation (11) multiplied by $f_m^{(r)}(z)$, the results subtracted and integrated between the limits a, b there results

$$\int_a^b f_k^{(s)}(z)\, f_m^{(r)}(z)\, dz = \frac{\left[f_k^{(s)}(z)\frac{df_m^{(r)}}{dz} - f_m^{(r)}(z)\frac{df_k^{(s)}}{dz}\right]_a^b}{k_o^2\left[n_s^2(a) - n_r^2(a) - (\beta_k^{(s)})^2 + (\beta_m^{(r)})^2\right]} \qquad (12)$$

To derive Equation (12) use has been made of the fact that $n_s^2(z) - n_r^2(z)$ is independent of z (the condition of equal gradients). By studying the limit $r \to s$, $m \to k$ it can also be shown that

$$\int_a^b (f_k^{(s)}(z))^2 dz = \frac{1}{k_o^2}\left[(k_o/|\alpha|)^{2/3}\, \alpha\, q(f_k^{(s)}(z))^2 + \frac{1}{\alpha}\left(\frac{df_k^{(s)}}{dz}\right)^2\right]_a^b \qquad (13)$$

where α is the gradient of n^2 in the layer being addressed and q is given by

$$q = (k_o/|\alpha|)^{2/3}(n^2(a) + \alpha\,(z - a) - (\beta_k^{(s)})^2) \qquad (14)$$

The contribution to the height gain integral from $-\infty$ to 0 is

$$\int_{-\infty}^{0} f_k^{(s)}(z)\, f_m^{(r)}(z)\, dz = f_k^{(s)}(o)\, f_m^{(r)}(o)/\left(ik_o(\sqrt{N_s^2 - (\beta_k^{(s)})^2} + \sqrt{N_r^2 - (\beta_m^{(r)})^2})\right) \qquad (15)$$

where

$$N_p = \sqrt{\frac{\sigma_p}{i\omega\epsilon_o} + \frac{\epsilon_p}{\epsilon_o}} \quad ; \quad I_m(N_p) < 0 \qquad (16)$$

σ_p = ground conductivity in slab p

ϵ_p = ground permittivity in slab p

ω = circular rf frequency

ϵ_o = free space permittivity

For comparisons made with the Cho and Wait [1978] results given in the following section, Equation (12) is all that is required for evaluation of the height gain integrals needed to evaluate the cumulative mode conversion coefficients given by Equations (7) and (9). For the evaporation duct studies the integrals are numerically evaluated between $z = 0$ and $z = z_u$. Above z_u the gradients of n^2 for all slabs are taken to be equal (118 M-units/km). The required height gain integrals then assume the form

$$\int_\Omega f_k^{(s)}(z)\, f_m^{(r)}(z) dz = \int_o^{z_u} f_k^{(s)}(z)\, f_m^{(r)}(z) dz - \frac{\left[f_k^{(s)}(z_u)\frac{df_m^{(r)}(z_u)}{dz} - f_m^{(r)}(z_u)\frac{df_k^{(s)}(z_u)}{dz}\right]}{k_o^2\left[n_s^2(z_u) - n_r^2(z_u) - (\beta_k^{(s)})^2 + (\beta_m^{(r)})^2\right]}$$

$$+ f_k^{(s)}(o)\, f_m^{(r)}(o)/\left(ik_o\left(\sqrt{N_s^2 - (\beta_k^{(s)})^2} + \sqrt{N_r^2 - (\beta_m^{(r)})^2}\right)\right) \qquad (17)$$

Although not required, the integral relating to the norm can be numerically evaluated as follows:

$$\int_\Omega (f_k^{(s)}(z))^2 dz = \int_o^{z_u} (f_k^{(s)}(z))^2 dz - \frac{1}{k_o^2}\left[(k_o/|\alpha_u|)^{2/3}\, \alpha_u q(z_u)(f_k^{(s)}(z_u))^2 + \frac{1}{\alpha_u}\left(\frac{df_k^{(s)}(z_u)}{dz}\right)^2\right]$$

$$+ (f_k^{(s)}(o))^2 / \left[2ik_o\sqrt{N_s^2 - (\beta_k^{(s)})^2}\right] \qquad (18)$$

where the gradient of n^2 above z_u is denoted by α_u and where

$$q(z_u) = (k_o/|\alpha_u|)^{2/3}\ (n^2(z_u) - (\beta_k^{(s)})^2) \qquad (19)$$

Equation (18) provides an excellent check on the numerical integration since the height gains can be normalized by using the result given in Equation (13).

SIGNAL LEVEL AND PATH LOSS

In terms of the cumulative mode conversion coefficients defined by Equations (7) and (9) and the normalized height gains discussed in Section II, the signal level relative to free space in slabs p≠1 is given by

$$E(dB)=10Log_{10}\left[\frac{2\pi x^2}{k_o^2 a \sin(x/a)}\ \left|\sum_m (k_o\beta_m^{(1)})^{1/2} \exp(-ik_o\beta_m^{(1)}x_2) f_m^{(1)}(z_T) \cdot \sum_j a_{jm}^{(p)} f_j^{(p)}(z_R) \exp(-ik_o\beta_j^{(p)}(x-x_p))\right|^2\right] \qquad (20)$$

When p = 1, E is simply

$$E(dB)=10Log_{10}\left[\frac{2\pi x^2}{k_o^2 a \sin(x/a)}\ \left|\sum_m (k_o\beta_m^{(1)})^{1/2}\ f_m^{(1)}(z_T) f_m^{(1)}(z_R) \exp(-ik_o\beta_m^{(1)}x)\right|^2\right] \qquad (21)$$

In these equations x is the range, z_T the transmitter height, z_R the receiver height, a is the earth's radius and all dimensions are mks. Sometimes signal levels expressed in terms of path loss are desired. In terms of signal level relative to free space, the path loss is given by

$$PL(dB) = 32.45 + 20Log_{10}(x) + 20Log_{10}(f) - E(dB) \qquad (22)$$

In Equation (22), x is in kilometers and the frequency, f, in MHz. In the following sections, results based on the above equations will be given for several laterally inhomogeneous tropospheric environments.

COMPARISONS WITH THE RESULTS OF CHO AND WAIT

In the paper by Cho and Wait [1978], a number of slab model mode conversion results associated with elevated layers at 200 MHz were presented. Figure 2 shows two slab models (c) and (d) used by Cho and Wait which are re-examined in this study and compared with the parabolic equation results of Ko and Burkom [1987]. Theoretical modified refractivity structure (apart from layer height) used in each slab is shown at the top of Figure 2. Also, the number of modes used in each slab at 200 MHz is indicated. In case (c), the layer height is 600 m out to a range of 200 km where the layer height then increases linearly to 1000 m at 500 km and then decreases linearly back to 600 m at a range of 800 km. Beyond 800 km the layer height remains uniform at 600 m. Case (d) is the inverse of (c) with the layer height of 1000 m invariant out to a range of 200 km. The layer height then decreases linearly to 600 m at 500 km and then increases to 1000 m at a range of 800 km. Beyond 800 km the layer height remains uniform at 1000 m. Cho and Wait approximate the linear variations with 30 km thick slabs.

Figures 3 and 4 show comparisons of the signal level variation with height between the current results and those of Cho and Wait for ranges of 350, 500, 650, 800 and 1000 km. The transmitter height for Figure 3 is 600 m and 1000 m for Figure 4. The excellent agreement between results illustrates their numerical equivalence. A point of considerable importance is that in replicating the results of Cho and Wait their exact slab thickness of 30 km has been used.

Slab convergence properties for cases (c) and (d) at a range of 350 km are shown in Figures (5) and (6) respectively. Shown are results for the four slab thicknesses of 30, 15, 7.5 and 3.75 km. It is evident that in both instances slab convergence is not approached until the slab thickness is reduced to at least 7.5 km and even then below about 200 m the convergence is questionable for case (c). Comparison of results for the 3.75 and 30 km slab thicknesses show a significant difference. These results are also representative of the convergence behavior at other ranges.

Comparisons between mode conversion results and parabolic equation results [Ko and Burkom, 1987] for the five ranges are given in Figures (7) and (8). Figure (7) applies to case (c) and Figure (8) to case (d). Shown are comparisons between the Cho and Wait results (slab thickness 30 km), the mode conversion results for a slab thickness of 3.75 km and the parabolic equation results. Generally, the parabolic equation results are in better agreement with the 3.75 km slab thickness mode conversion results, though the differences are larger than one might have hoped. This is particularly true for case (c) at the ranges of 650 and 800 km where the deep, broad nulls above 1000 m do not show up in either the parabolic equation calculations nor in the mode conversion calculations with the 30 km slab thickness. The reason for the disparity is not known. Differences between parabolic equation results and mode conversion (3.75 km slab) calculations below about 400 m are not surprising since comparisons (not discussed here) with

extended mode calculations indicate higher order modes can affect the rather weak signal levels there. Comparisons for both cases (c) and (d) between the 3.75 km slab mode conversion result and the parabolic equation result at the furthest distance of 1000 km is quite good. The reason for this is probably that higher order modes are much attenuated in travelling the 200 km distance beyond the end of the lateral inhomogeneity.

EVAPORATION DUCT RESULTS

The evaporation duct is a nearly permanent, shallow ($\leq$ 30 m) surface duct created by the rapid decrease of moisture immediately above the ocean surface. Air in contact with the sea is saturated with water vapor which decreases with height until an ambient value is reached. This decrease of water vapor with altitude causes the index of refraction to decrease near the sea surface, and this in turn produces a duct within which grazing rays are bent towards the earth. Figure 9 shows a schematic of the behavior of the modified refractive index, m, for such a duct. The most important duct parameter is the duct height, h_d, where the m profile minimizes. Below h_d the m profile varies nearly logarithmically. Above, h_d, the variation is roughly linear. Because these ducts are vertically thin, strong trapping infrequently occurs below about 3 GHz.

Waveguide and mode conversion results are presented in this section for several evaporation duct environments and are compared with corresponding parabolic equation results of Ko and Burkom [1987]. In all cases the transmitter is at 5.0 m above sea level, the frequency is 9.6 GHz, the polarization is horizontal and a smooth sea surface is assumed. The evaporation duct profiles, approximated by linear segmentation, were supplied by H. Hitney. They were obtained from numerical solutions to the equations for calculating the duct height from meteorological parameters [Jeske, 1973] subject to the assumption of neutral atmospheric stability.

Ducting environments considered in the remainder of this section are:

(1) Horizontally uniform guides characterized by the standard atmosphere (i.e. linear layer with gradient of 118 M-units/km), as well as evaporation ducts characterized by 2 m, 6 m, 8 m and 16 m duct heights.

(2) A 6 m duct height at the beginning of a 35.2 km path increasing lineraly to 8 m at mid-path then decreasing linearly to 6 m at the end of the path (i.e., 6 m at 0 km, 8 m at 17.6 m and 6 m at 35.2 km).

(3) A 2 m duct height at the beginning of a 35.2 km path increasing linearly to 8 m at mid-path then decreasing linearly to 2 m at the end of the path (i.e., 2 m at 0 km, 8 m at 17.6 km and 2 m at 35.2 km).

(4) A 2 m duct height at the beginning of a 35.2 km path increasing linearly to 16 m at the end of the path (i.e. 2 m at 0 km and 16 m at 35.2 km).

Figure 10 gives waveguide (WG) and parabolic equation (PE) height gain results at 35.2 km for uniform guides characterized by the standard atmosphere as well as evaporation ducts with duct heights of 8 and 16 meters. The horizontal axis is signal level in dB relative to free space and the vertical axis is altitude in meters. In all cases the waveguide and parabolic equation results are in excellent agreement.

Figure 11 gives waveguide and parabolic equation range results for the same environments used for Figure 9. Results are in terms of path loss in this instance and the free space path loss is included for reference. Both transmitter and receiver height is 5 m. As before, the waveguide and parabolic equation results are in excellent agreement.

Figures 12 and 13 show mode conversion (MC) and parabolic equation results for case (2) enumerated above (i.e., the 6-8-6 m duct height variation). Height gain behavior is shown in Figure 12 at 35.2 km and range behavior for a receiver height of 5 m is shown in Figure 13. Also shown on the plots are waveguide results for the standard atmosphere and for evaporation ducts with 6 m and 8 m duct heights. As determined by convergence studies, 12 modes and 21 slabs were used for the mode conversion calculations. The latter are, as expected, in all instances bounded by the 6 m and 8 m waveguide results. Comparison between mode conversion and parabolic equation results is very good.

Comparisons between mode conversion and parabolic equation results for case (3) enumerated above (i.e., the 2-8-2 m duct height variation) are shown in Figures 14 and 15. Height gain behavior is shown in Figure 14 at 35.2 km and range behavior for a receiver height of 5 m is shown in Figure 15. Also shown on the plots are waveguide results for the standard atmosphere and for evaporation ducts with duct heights of 2 and 8 m. For the mode conversion calculations, the number of modes varied from 6 to 12 between the 2 and 8 m duct heights respectively and the lateral inhomogeneity was modeled with 121 slabs. Mode conversion and parabolic equation results are in good agreement and fall as expected between the 2 and 8 m duct height results. Some wiggles will be seen in the mode conversion results shown on Figure 15 in the neighborhood of 31 km. Those are due to slight discontinuities across slab boundaries.

Comparisons between mode conversion and parabolic equation results for case (4) enumerated above (i.e., the 2-16 m duct height variation) are shown in Figures 16 and 17. Height gain behavior is shown in Figure 16 at 35.2 km and range behavior for a receiver height of 5 m is shown in Figure 17. Also shown on the plots are waveguide results for the standard atmosphere and for evaporation ducts with duct heights of 2 and 16 m. For the mode conversion calculations, the number of modes varied from 6 to 12 between the 2 and 8 m duct heights respectively and 12 modes were also used for all duct heights greater than 8 m. Seventy one slabs were used to model the lateral inhomogeneity. Although quite small, discontinuities at slab boundaries will be seen on Figure 17. It is curious that the region just beyond 30 km where the most pronounced discontinuities occur is also the region where the parabolic equation results are most variable. That is probably a fortuitous occurrence. Overall, the mode conversion and parabolic equation results agree quite well.

CONCLUSIONS

In contrast with the 30 km thick slabs used by Cho and Wait, slab thicknesses of 3.75 km have been used in the present study. This appears to be adequate for the mode set used by Cho and Wait. Additional modes can influence height gain results below about 400 m. However, above that height it is believed that the original mode set along with the slab thickness of 3.75 km gives convergent results.

Though in most instances, comparison between the parabolic equation and mode conversion results for the Cho and Wait case is reasonable, there are two cases which show surprising differences above about 1000 m. They are the results for ranges of 650 and 800 km for case (c). Mode conversion results show broad deep nulls in the neighborhood of 1200 m which are more than 20 dB below the parabolic equation results. Reasons for the discrepancy are not known.

Comparisons between mode conversion and parabolic equation calculations for the laterally homogeneous and inhomogeneous evaporation duct environments considered are all very good.

Three major problems; slab size, number of modes and height gain integral evaluations beset mode conversion calculations. Even for the limited class of profiles considered in this study, the waveguide and mode conversion runs typically required hundreds of minutes of computer processing time whereas the parabolic equation results required minutes. Thus, the comparisons mentioned above, along with the ease of implementation and speed of the parabolic equation calculations relative to the mode conversion analysis, strongly supports the superiority of the parabolic equation method for systems applications involving laterally inhomogeneous tropospheric layering. It may even turn out that the most useful Navy application of parabolic equation solvers in the area of electromagnetics will be for propagation calculations in laterally homogeneous ducting environments.

REFERENCES

1. Baumgartner, G.B., Jr., XWGD: A computer program for trilinear tropospheric ducts, Naval Ocean Systems Center, Tech. Doc. 610, June 1983.

2. Budden, K.G., The Waveguide Mode Theory of Wave Propagation, London, England, Logus Press, 1961.

3. Cho, S.H. and J.R. Wait, Analysis of microwave ducts in an inhomogeneous troposphere, Pure Appl. Geophys., vol 116, pp 1118-1142, 1978.

4. Dockery, G.D. and G.C. Konstanzer, Recent advances in prediction of tropospheric propagation using the parabolic equation, Johns Hopkins APL Tech. Dig. 8, 404-412, 1987.

5. Hitney, H.V., J.H. Richter, R.A. Pappert, K.D. Anderson and G.B. Baumgartner, Jr., Tropospheric radio propagation assessment, Proc. IEEE, vol. 73, no. 2, pp 265-283, 1985.

6. Jeske, H., Die Ausbreitung elektromagnetischer Wellen im cm-bis m-Band uber dem Meer unter besono rer Berucksichtigung der meteorologischen Bedingungen in der maritimen Grenzschicht, Hamburger Geophys. Einzelschriften, Hamburg, De Gruyter and Co., 1965.

7. Kerr, D.E., Ed., Propagation of Short Radio Waves, New York, McGraw-Hill, 1951.

8. Ko, H.W. and H.S. Burkom, An analysis of EMPE code performance in a selection of laterally inhomogeneous ducting environments, Johns Hopkins APL, STD-M-515, June 1987.

9. Ko, H.W., H.S. Burkom, J.P. Skura and D.A. Roberts, An analysis of EMPE code performance in a selection of laterally inhomogeneous atmospheric-duct environments, Johns Hopkins APL Tech. Dig. 9, 89-100, 1988.

10. Ko, H.W., J.W. Sari and J.P. Skura, Anamalous microwave propagation through atmospheric ducts, Johns Hopkins APL Tech. Dig. 4, 12-26 (1983).

11. Midler, M.D., Rays and wave invariants for SOFAR channel propagation, vol 46, no. 5 (part 2), 1259-1263, 1969.

12. Morfitt, D.G. and C.H. Shellman, 'MODESRCH' an improved computer program for obtaining ELF/VLF/LF mode constants in an earth ionosphere waveguide, Interim Rep. 77T, prepared for the Defense Nuclear Agency by the Naval Electronics Center (Now Naval Ocean Systems Center), October 1976.

13. Pappert, R.A. and C.L. Goodhart, Case studies of beyond-the-horizon propagation in tropospheric ducting environments, Radio Sci., vol 12, no. 1, pp 75-87, 1977.

14. Tappert, F.D., The Parabolic Approximation Method, in Wave Propagation and Underwater Acoustics, J.B. Keller and J.S. Papadakis, eds., Springer-Verlag, pp 2224-286, 1971.

15. Wait, J.R., Electromagnetic Waves in Stratified Media, 2nd ed., New York: Pergamon, 1970.

16. Wait, J.R., Coupled mode analysis for a non-uniform tropospheric waveguide, Radio Sci., vol 15, pp 667-673, 1980.

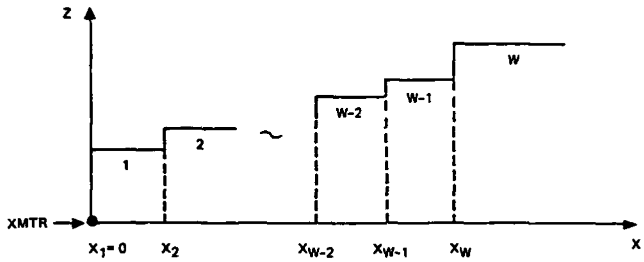


Figure 1. Slab Mode Conversion Model

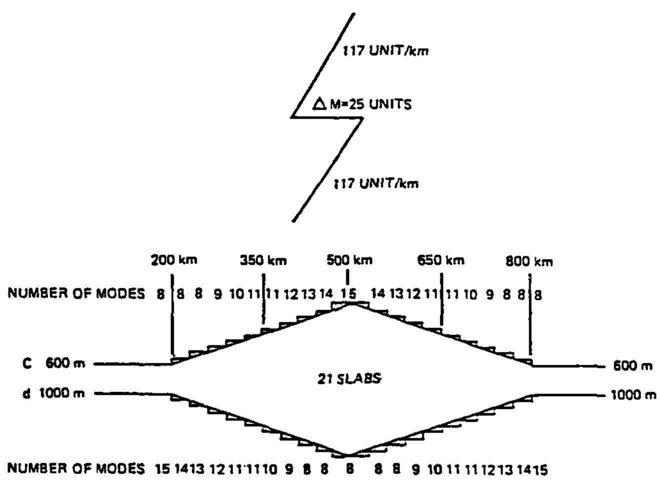


Figure 2. Slab Models for Comparison with Cho and Wait

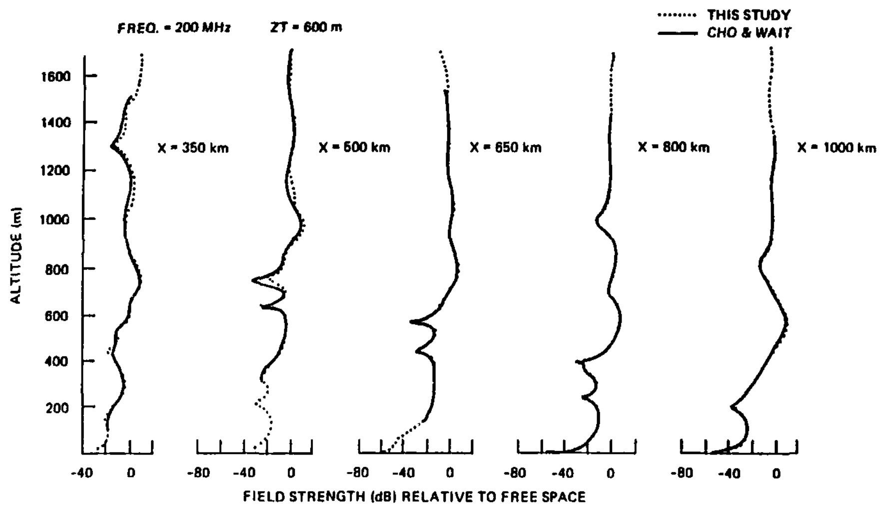


Figure 3. Height Gain Comparisons with Cho and Wait for Case (c)

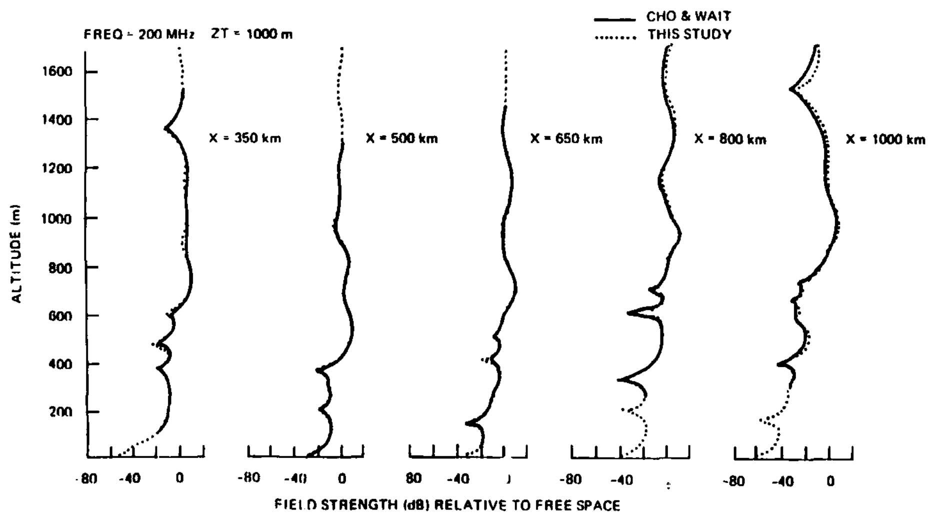


Figure 4. Height Gain Comparisons with Cho and Wait for Case (d)

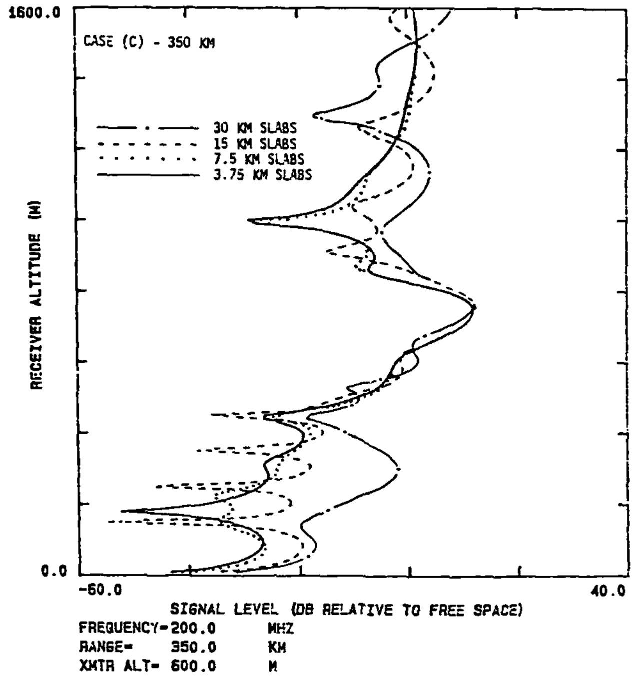


Figure 5. Slab convergence properites at 350 km for case (c).

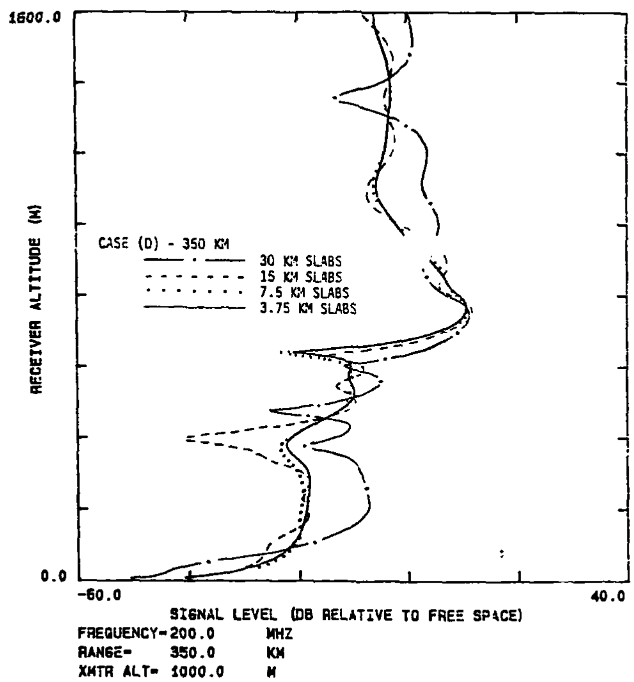


Figure 6. Slab convergence properties at 350 km for case (d).

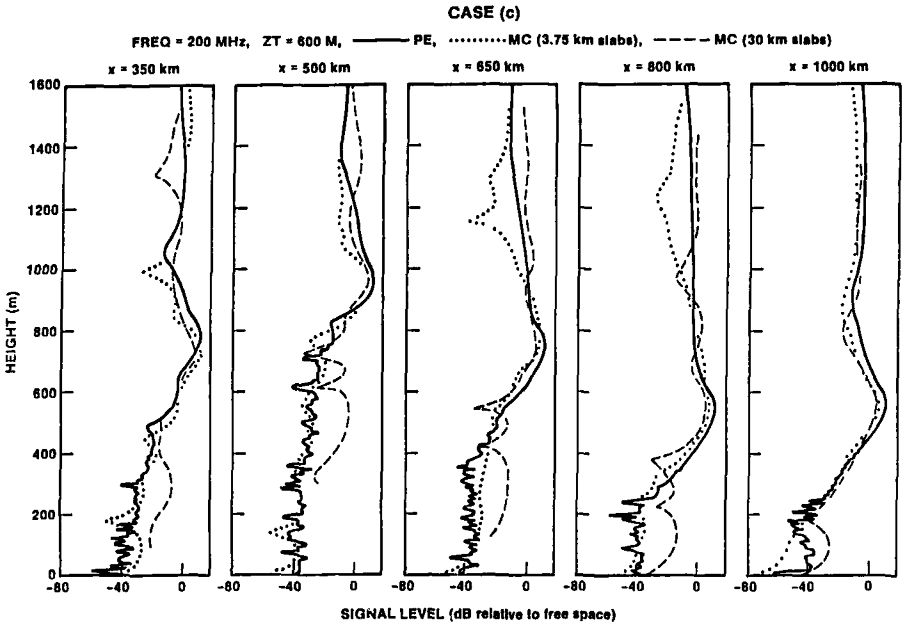


Figure 7. Mode Conversion (MC) and Parabolic Equation (PE) Field Strength Comparisons for Case (c).

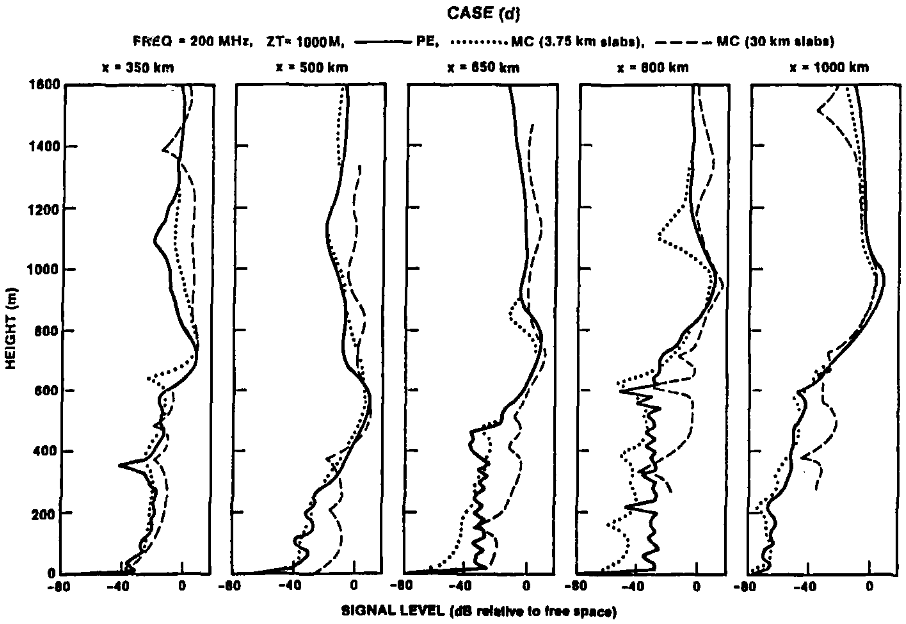


Figure 8. Mode Conversion (MC) and Parabolic Equation (PE) Field Strength Comparisons for Case (d).

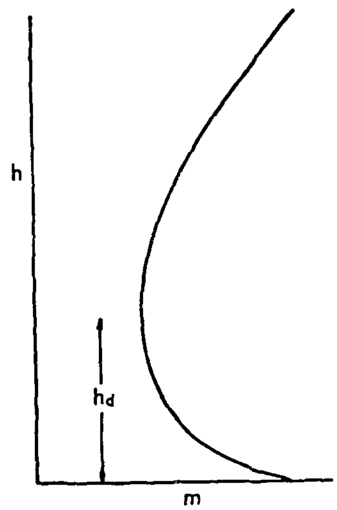


Figure 9. Evaporation Duct Schematic.

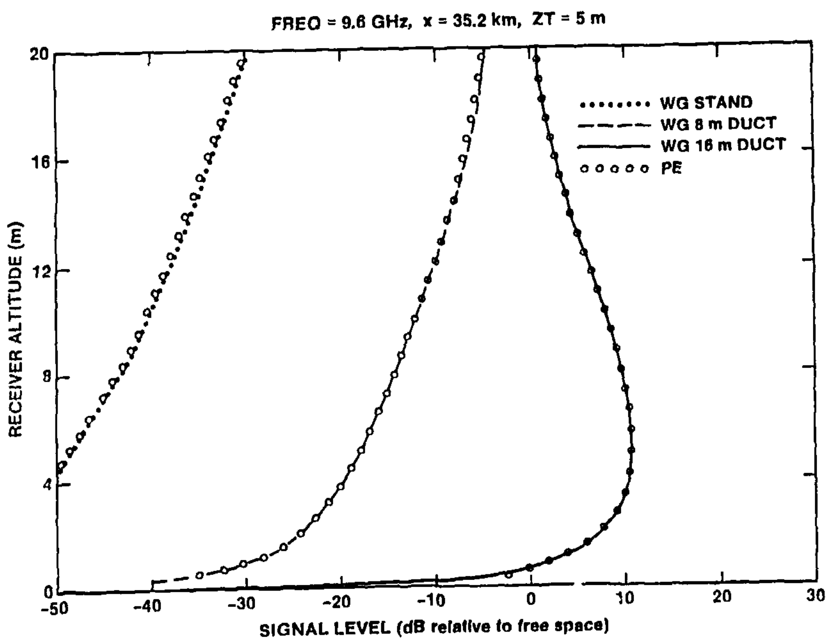


Figure 10. Waveguide (WG) and Parabolic Equation (PE) Height Gain Comparisons.

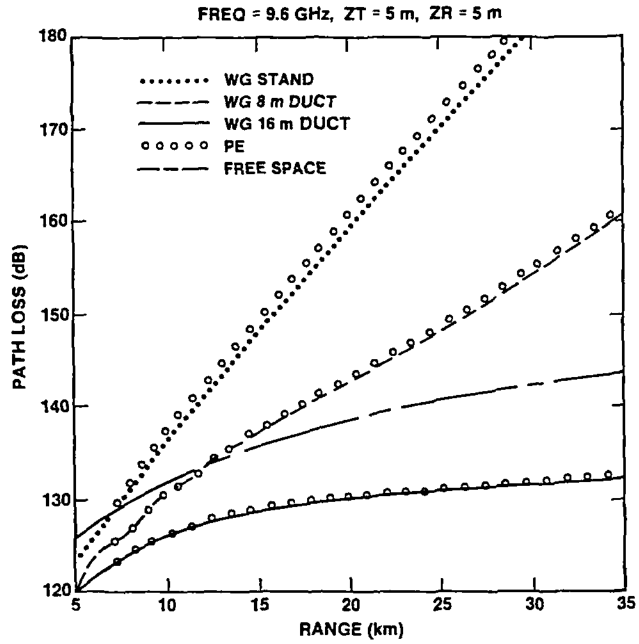


Figure 11. Waveguide (WG) and Parabolic Equation (PE) Range Comparisons.

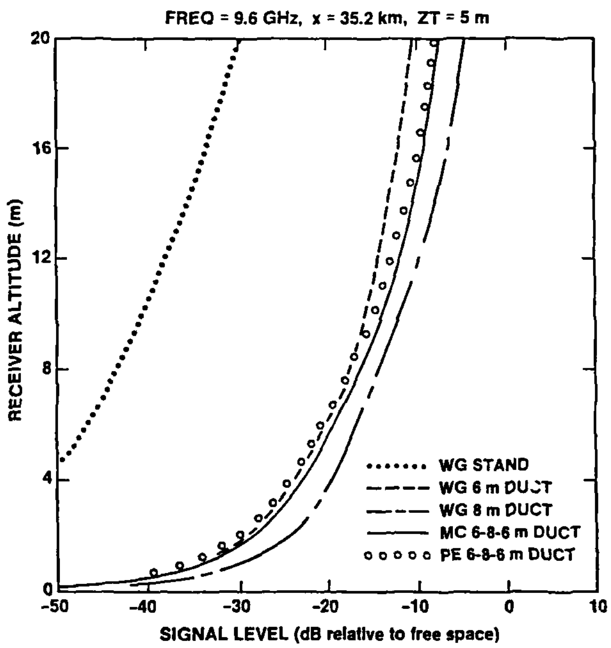


Figure 12. MC and PE Height Gain Comparisons for 6-8-6m Duct and WG Reference Plots.

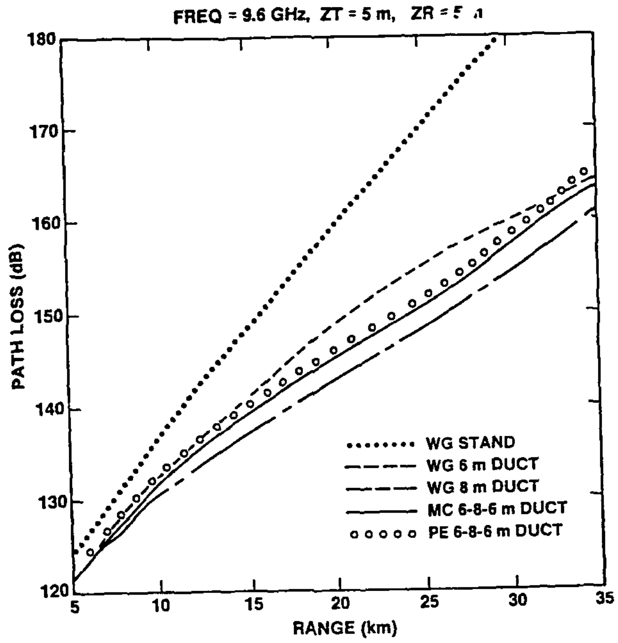


Figure 13. MC and PE Range Comparisons for 6-8-6m Duct and WG Reference Plots.

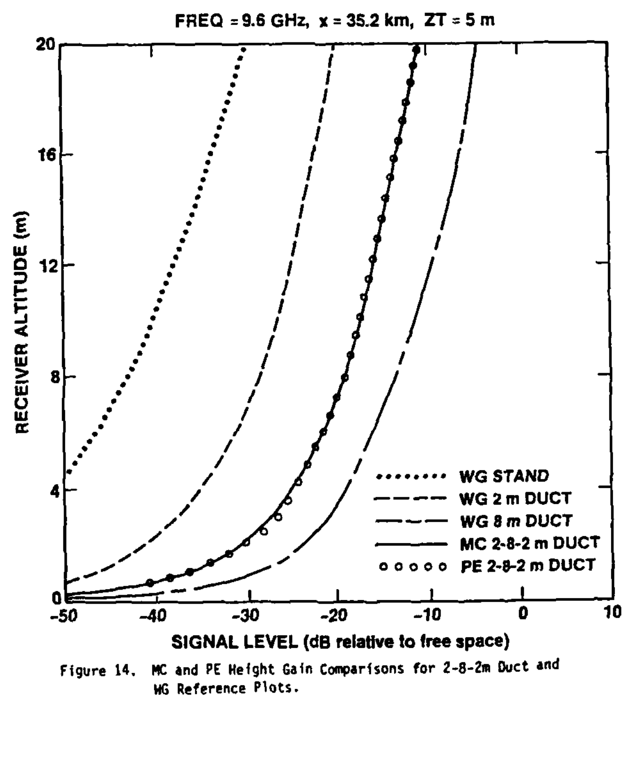


Figure 14. MC and PE Height Gain Comparisons for 2-8-2m Duct and WG Reference Plots.

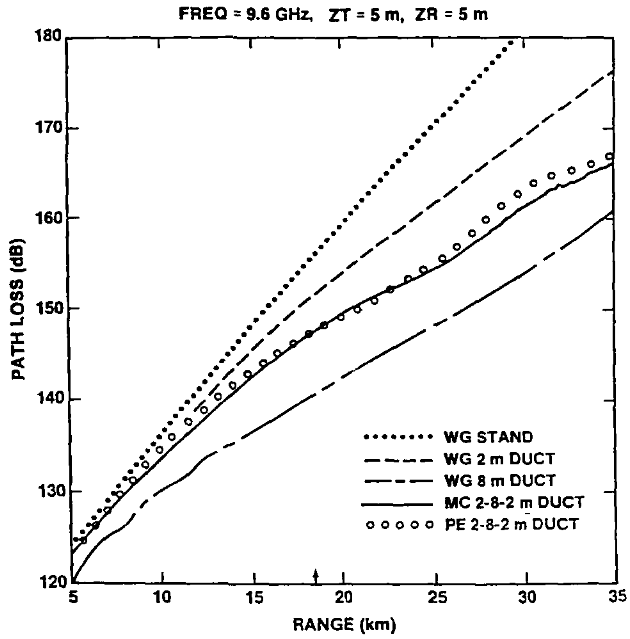


Figure 15. MC and PE Range Comparisons for 2-8-2m Duct and WC Reference Plots.

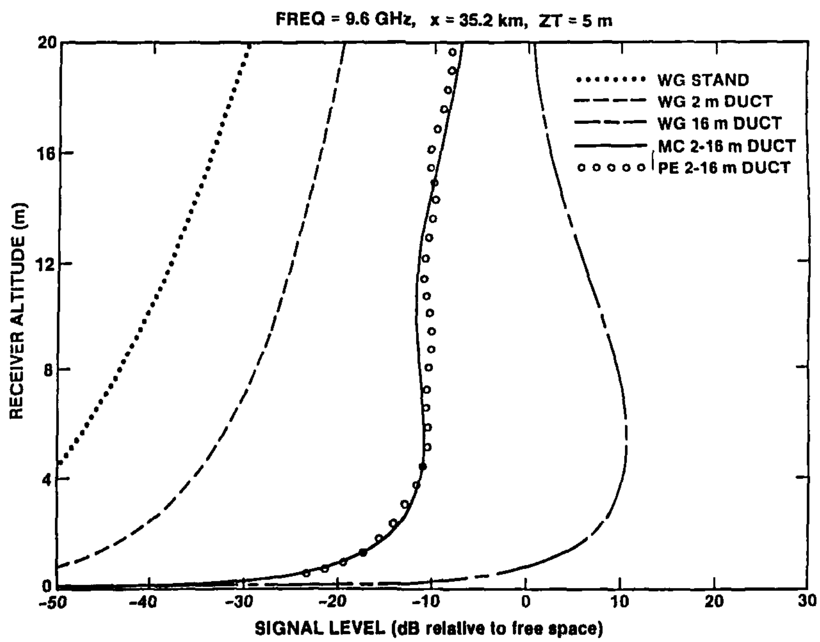


Figure 16. MC and PC Height Gain Comparisons for 2-16m Duct and WG Reference Plots.

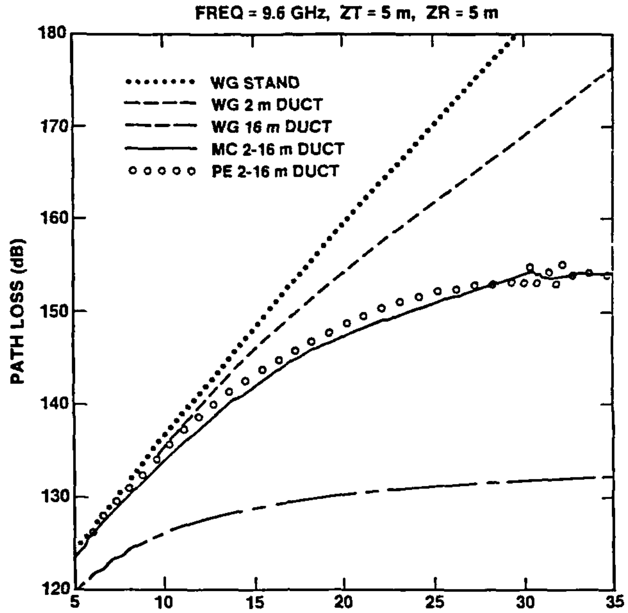


Figure 17. MC and PE Range Comparisons for 2-16m Duct and WG Reference Plots.

DISCUSSION

G. DOCKERY, US

Was an infinite refractive index gradient used for the calculations corresponding to the Cho and Wait study? If so, refractive index gradient limits may have been violated in the parabolic equation calculations.

AUTHOR'S REPLY

No. A gradient of about -2500 M units/m was used to model the abrupt layer and a criterion given by Ko, et al. (reference 9) indicates that such gradients should be manageable with their PE solver.

L. FELSEN, US

Would the adiabatic modes work for some (which?) ... r range-dependent profiles, thereby not requiring the full coupled mode algorit...?

AUTHOR'S REPLY

Since only mode conversion calculations were performed, I can only speculate about applicability of the adiabatic method. Results for the laterally inhomogeneous evaporation ducts suggest that the adiabatic approximation may be adequate for those profiles. However, based on the magnitudes of the cumulative mode conversion coefficients, I would not expect that to be the case for the range dependent elevated ducts.

PROPAGATION PREDICTION FOR THE NORTH SEA ENVIRONMENT

by

R.Vieth
Dornier System GmbH
Management Consulting
Postfach 1360
D-7990 Friedrichshafen
Federal Republic of Germany

SUMMARY

Evaporation ducting can have an important influence on the propagation of electromagnetic waves. The dependence of those ducting conditions on geographic location requires an estimate of occurrence and effects of ducting in various areas. Duct height statistics using long term statistical meteorological data in combination with propagation models are used for this purpose. In this paper, Jeske's propagation measurements [1] during 1961 at the German coast of the North Sea were taken and compared with the calculated results from the combination of the statistical weather data base and the propagation models, as well as another measurement program performed in Greece [2]. A brief description of the models is followed by an example of the results of the Greek measurements. The German experimental data and duct height distributions for that region are described. Finally the results of measurements and calculations are discussed. A good agreement was found between measured propagation data and predictions based on climatological averages.

INTRODUCTION

The evaporation duct is known as a frequently occurring propagation mechanism influencing radar propagation near the sea surface. This propagation effect has to be taken into account when predicting the performance of radar systems, especially for shipboard applications. Duct height statistics and propagation models are used for evaluation of radar systems as well as electromagnetic tactical decision aids software. Measurement of both, meterological data and radio signal levels, in California and Florida (USA) and the Aegean Sea (Greece) have been performed to validate the models. It has been shown that existing models together with the weather data base are able to predict propagation effects for these climatological regions. In many areas of interest to NATO there are quite different climates and therefore different duct height distributions. Here the models are used to predict the results of a long term measurement in the North Sea environment.

MODELS

The propagation model MLAYER was developed at the Naval Ocean Systems Center by Baumgartner and later modified by Pappert and is described in reference [3]. MLAYER can solve the modal equation for an arbitrary vertical multiple-linear-segment refractivity profile using a root finding scheme that guarantees that all modes with an attenuation rate less than a given value are found. Surface roughness is taken into account through a modification to the surface reflection coefficient based on the variance of surface heights, which in turn is calculated from surface wind speed. For all results shown in this paper, surface roughness was accounted for by setting the surface wind speed to 7 m/sec, which corresponds to world average conditions. Horizontal homogeneity of refractivity conditions is assumed by this model.

Jeske [1] and Paulus [4] have shown that evaporation duct height can be determined as a function of air- and sea temperature, relative humidity and wind speed. This method has been used to compute evaporation duct height frequency distributions for 292 oceanic areas, defined by 10 degree latitude by 10 degree longitude Marsden Squares, using a 15 year subset of marine surface observations compiled by the United States National Climatic Data Center, Ashville, North Carolina. This data base, known as Surface Duct Summary (SDS), is part of the Engineer's Refractive Effects Prediction System (EREPS) [5].

The propagation model PROPR contained in EREPS accounts for multipath, diffraction, tropospheric scatter, surface based ducts, evaporation ducts and water vapor absorbtion. In case of an evaporation duct, the path loss near and beyond the horizon is approximated with easy-to-calculate algorithms, which use a single mode model to determine path loss as a function of duct height. Therefore the results from PROPR are approximations of the full waveguide solution given by MLAYER for the region under investigation.

Path loss for the given terminal heights and path length versus evaporation duct heights from 0 to 40 m in 2 m steps were calculated with both propagation models. These results were then weighted by the annual percent of occurrence of evaporation duct height in each 2 m interval for the Marsden Square within which the experiment was performed to give accumulated frequency distributions of path loss, which are then compared to the measured distribution.

AEGEAN SEA EXPERIMENT

During 1972 a series of propagation measurements were made over a distance of 35.2 km between the Greek islands of Naxos and Mykonos in the Aegean Sea. The transmitters for 3.0 and 9.6 GHz were installed on Naxos at 4.8 m above mean sea level. Receivers were located at Mykonos at 19.2 m. Standard-atmosphere radio horizon for this geometry is about 27 km.

Path loss was measured during four two-to-three week periods in February, April, August and November. All recorded data were averaged over a 5 minute period, with samples taken every 15 minutes, 24 hours a day.

Figure 1 shows the annual duct height distribution from EREPS for the area in which the measurements were made. Note that the most common duct heights are between 10 and 12 m, with duct heights greater than 30 m being quite rare. The calculated path loss combined with this distribution gives the accumulated frequency distribution of path loss for the two frequencies in Figures 2 and 3. Modeled distributions are shown by solid curves (MLAYER) and dashed curves (EREPS), whereas the dotted curves are the measured path loss distributions. Reference lines are included showing the free-space and diffraction-field path loss values for each frequency. The diffraction field was calculated based on a standard 4/3 effective-earth-radius atmosphere. Figure 2 shows a substantial decrease in path loss compared to diffraction much of the time at 3.0 GHz, with the median path loss being 15 dB less than diffraction. The modeled and measured distributions are in excellent agreement at the higher path loss values, but differ at the lower path loss levels. It is speculated that effects from surface based ducts created by elevated trapping layers are responsible for this discrepancy, which have not been accounted for in the modeling. The results for 9.6 GHz in Figure 5 show the modeled and measured curves in good agreement over most of the range, with median values showing some 26 to 28 dB reductions compared to diffraction. At this frequency, the median values are much closer to the free-space value than to the diffraction value. As with the 3.0 GHz case, there is a discrepancy between measured and modeled results at the lower path loss values.

NORTH SEA EXPERIMENT

Jeske used a one way propagation path between Weddewarden (near Bremerhaven at the north coast of Germany) and the island Helgoland (German Bight) for field strength measurements at 0.6, 2.3, 6.8 GHz. The distance between the transmitters and receivers was 77.2 km. The transmitter antennas on the coast were located at 35, 28 and 29 m, and the receiver antennas were placed on the island at heights of 32, 31 and 33 m, respectively. The radio horizon for standard atmosphere for this geometry is about 46 km. Jeske's field strength measurements were obtained from continuously plotted receiver input voltages. Hourly median, as well as the maximum and minimum, field strength values were taken from these plots. To compare these field strength results with the modeled data, field strength was converted to path loss.

The annual duct height distribution from EREPS for the North Sea area is shown in Figure 4. Note by comparing Figures 1 and 4, that evaporation duct heights are typically less in the North Sea than in the Aegean. Figures 5 through 7 show the percentage of time the path loss exceeds the abscissa value for each frequency. Comparisons of the measured data and the modeled results are presented in the same format as the Aegean results. The path loss reduction relative to the diffraction level (i,e. signal improvement over diffraction) increases with increasing frequency as in the Aegean results. Here the signal improvement of the one way path at the 50 % level is measured to be 6, 12 and 33 dB above the diffraction path loss. The MLAYER model gives 3, 13 and 34 dB and the EREPS calculations 4, 10 and 34 dB, respectively. For the 0.6 GHz case the difference between measured and MLAYER data is 3 dB, EREPS shows an error of 2 dB. The difference for 2.3 and 6.8 GHz lies below 2 dB at the 50 % level. As with the Aegean results the overall comparisons between the measured and modeled results for the North Sea experiment are considered to be quite good.

DISCUSSION

The frequent occurrence of very low path loss values in the measured data is believed to be due to either surface based ducts or super-refractive effects. The models used here only account for evaporation duct effects. Surface-based ducts and super-refractive effects were not modeled. Surface-based ducts occur in the Aegean area about 11 % of the time annually and in the North Sea area about 1.7 % of the time. Note that the difference between the measured and the calculated curves for the lower path loss values is less for the North Sea results than for the Aegean results, perhaps due to the lesser occurrence of surface-based ducts. As radio frequency increases for both experiments, the difference between the modeled and measured curves for low path loss values generally decreases, because lower frequencies are less sensitive to evaporation duct effects than higher frequencies. At lower frequencies, surface based ducts and super-refractive effects seem to be the more important propagation mechanisms and are thus responsible for the lower path loss values.

There is also a difference between the measured and calculated curves at high path loss values. The measured path loss exceeds the standard atmosphere diffraction level when rain or sub-refractive conditions occur. The duct heights in the EREPS data base are set to 0 for sub-refractive conditions, which implies standard atmosphere conditions. Therefore the calculated values can't exceed the diffraction level. Jeske made meteorological measurements during his experiment from a ship stationed approximately at mid-path, that show sub-refractive conditions occurring about 3 % of the time. Had this effect been included in the models, it is thought that the high-path-loss comparisons would be better.

Despite the differences at the lower and upper ends of the curves, modeled and measured data are in very good agreement. A perfect match of curves should not be expected, because of the variation between the 15-year evaporation duct data base used and the actual duct height distribution in 1961 or 1972 in the area of the experiments. However, a favorable comparison has been made of the duct height distribution from Jeske's meteorological measurements and the long-term distribution shown in Figure 4. Noteworthy of this study

is the substantial reduction of path loss, or increase in signal, that is both modeled and measured on over-the-horizon paths, particulary at higher frequencies. Many system engineers use standard diffraction theory to estimate required margins or to otherwise specify their systems. Especially at higher frequencies, the median levels may actually be closer to free space than to diffraction, for a net gain of 20-30 dB or more. At frequencies as low as 2 GHz, there is also a substantial increase. For the radar application, this effect could give median detection ranges far in excess of those predicted on basis of diffraction alone.

CONCLUSION

The comparison of measured and calculated path loss distributions showed very good agreement at frequencies where the evaporation duct is dominant. Even though the North Sea region has a lower average evaporation duct height than the Aegean Sea, modeled and measured data match better, which should be expected due to the lesser occurrence of surface based ducts. It was possible to predict the measured path loss distribution for the North Sea experiment within a few dB by calculations using the MLAYER and EREPS propagation models and the evaporation duct height distribution from the SDS data base, except for the extreme upper and lower path loss values.

REFERENCES

[1] H.Jeske,"Die Ausbreitung elektromagnetischer Wellen im cm- bis m-Band über dem Meer unter besonderer Berucksichtigung der Bedingungen in der maritimen Grenzschicht", 1965, Hamburger Geophysikalische Einzelschriften

[2] J.H.Richter, H.V.Hitney, "Antenna Heights for the Optimum Utilization of the Evaporation Duct, Vol. 1 and 2", January 1988, Naval Ocean Systems Center Technical Document 1209

[3] H.V.Hitney, J.H.Richter, R.A.Pappert, K.D.Anderson, G.B.Baumgartner, "Tropospheric Radio Wave Propagation", Proc. IEEE, vol. 73, no. 2, February 1985

[4] R.A.Paulus, "Practical Application of an Evaporation Duct Model", Radio Sci., vol. 20, no. 4, July-August 1985, pp 887-896

[5] H.V.Hitney, A.E.Barrios,G.E.Lindem, "Engineer's Refractive Effects Prediction System (EREPS) Revision 1.0 Users Manual", July 1988, Naval Ocean Systems Center Technical Document 1342

ACKNOWLEDGEMENTS

I wish to thank H.V.Hitney for many valuable contributions and his help in preparing this paper.

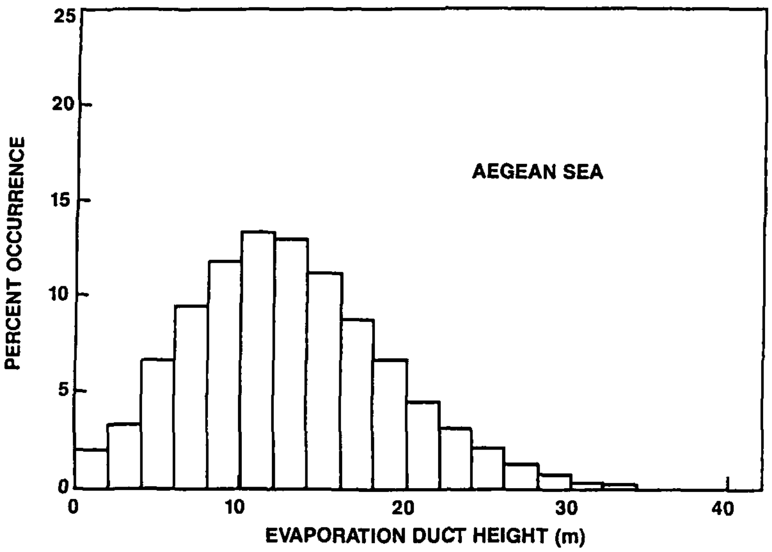


Figure 1 Histogram of evaporation duct height occurrence in the Aegean Sea

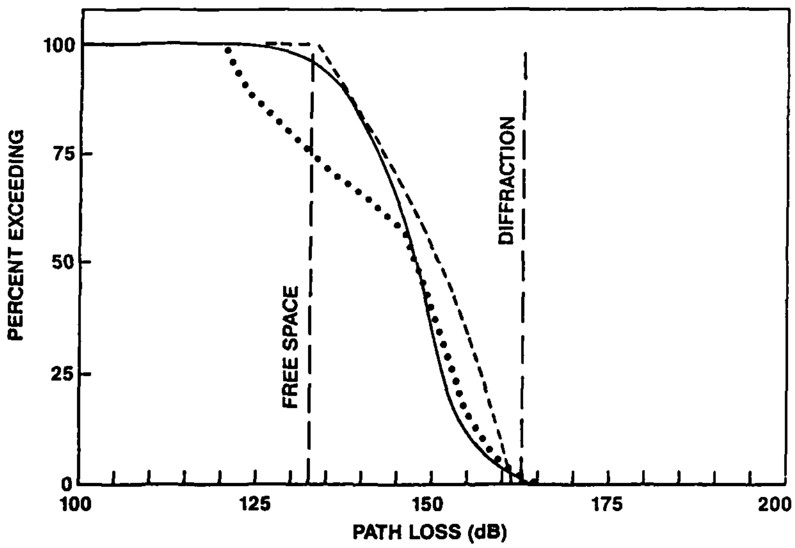


Figure 2 Path loss distribution for the Aegean Sea experiment at 3.0 GHz
•••••••• measured ———— MLAYER - - - - - - EREPS

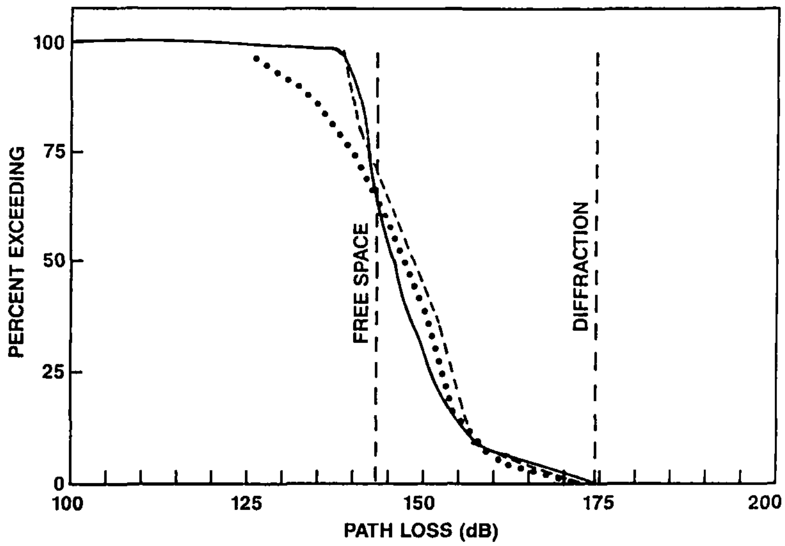


Figure 3 Path loss distribution for the Aegean Sea experiment at 9.6 GHz
•••••••• measured ———— MLAYER - - - - - - EREPS

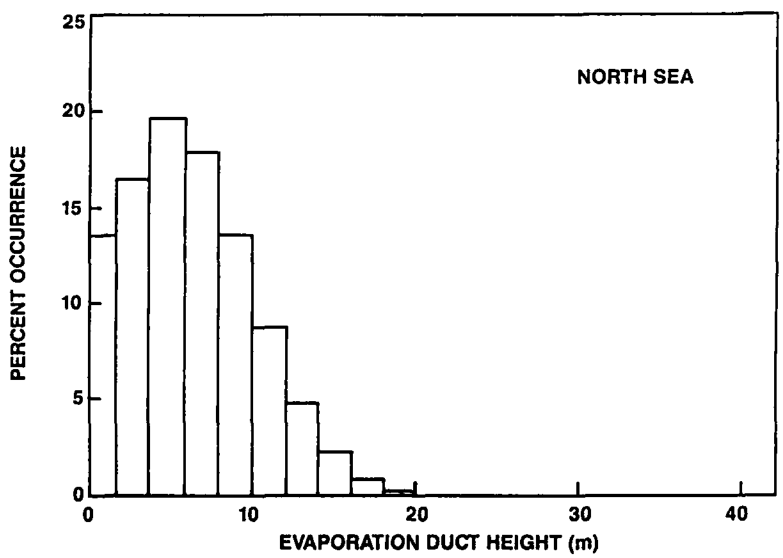


Figure 4 Histogram of evaporation duct height occurrence in the North Sea

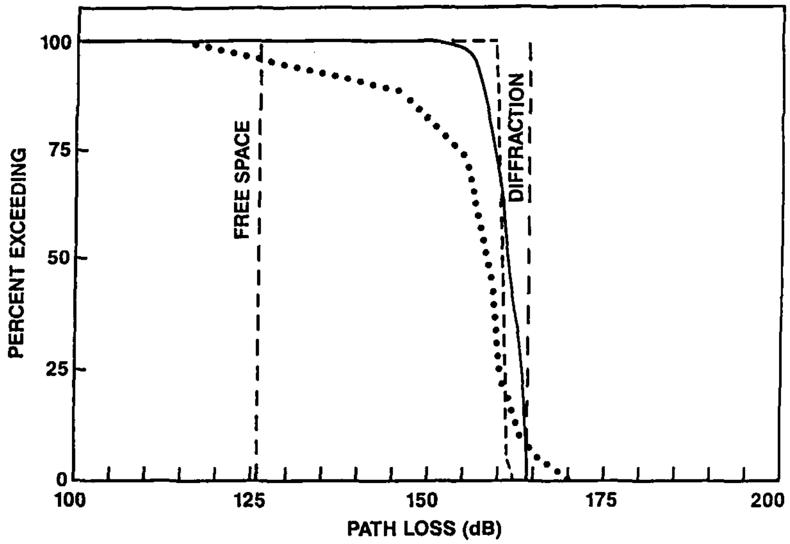


Figure 5 Path loss distribution for the North Sea experiment at 0.6 GHz
•••••••• measured ———— MLAYER - - - - - - EREPS

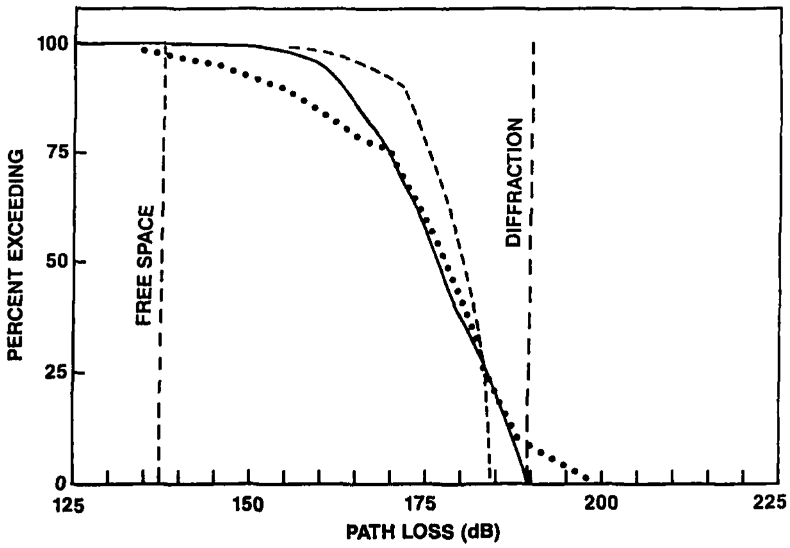


Figure 6 Path loss distribution for the North Sea experiment at 2.3 GHz
•••••••• measured ———— MLAYER - - - - - - EREPS

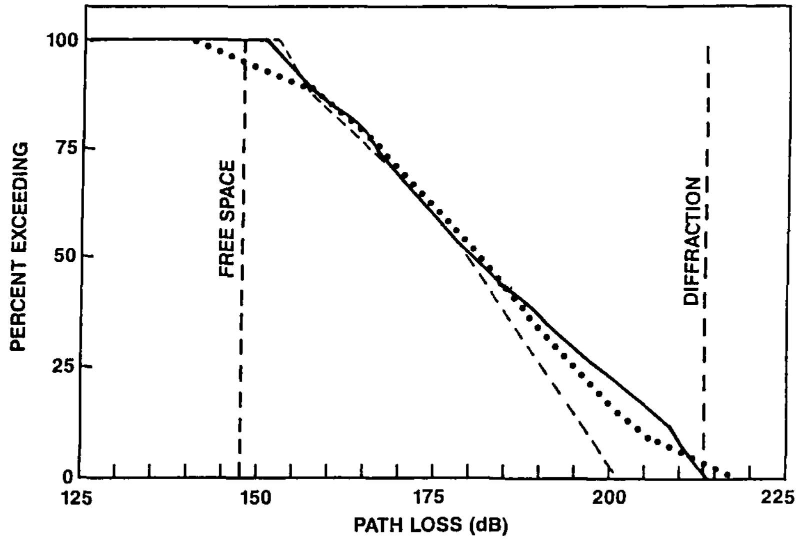


Figure 7 Path loss distribution for the North Sea experiment at 6.8 GHz
•••••••• measured ———— MLAYER - - - - - - EREPS

DISCUSSION

F. CHRISTOPHE, FR

Could the equipment you described be used for the analysis of short term amplitude fluctuations, and of differential phase?

AUTHOR'S REPLY

The models used are propagation models and long term statistical duct height distributions. Therefore the analysis of short-term amplitude fluctuations and differential phase was not possible. The measured data of the German experiment however, contains fading information.

ASSESSMENT OF ANOMALOUS PROPAGATION PREDICTIONS USING MINISONDE REFRACTIVITY DATA AND THE PARABOLIC EQUATION METHOD

by

M.F. Levy and K.H. Craig
Radio Communications Research Unit
Rutherford Appleton Laboratory
Chilton, Didcot, OX11 0QX
United Kingdom

SUMMARY

Accurate and easy-to-use systems are now available to measure the refractive index structure in the troposphere. Combined with an efficient package for solving Maxwell's equations for radiowave propagation based on the Parabolic Equation Method, they provide a powerful tool for real time forecasts of anomalous propagation. Quantitative field-strength predictions can be obtained in a few minutes from measured meteorological profiles, on a desk-top computer. However, measurement errors inevitably lead to variability in the predictions, which should be seen in a stochastic light. Some examples based on experimental data are presented, at X-band and in the millimetre wave part of the spectrum, as well as a sensitivity analysis using statistical simulations. The Parabolic Equation Method copes well with noisy refractivity data, and the predictions are fairly insensitive to radiosonde errors.

1. INTRODUCTION

The problem of providing quantitative predictions for microwave propagation in the troposphere has, until recently, been extremely difficult to tackle. There were difficulties in obtaining meteorological data, because of the poor resolution of the radiosonde network, and the available data were inadequate for assessing refractivity effects. The large time constant of the humidity sensor resulted in the smoothing of humidity gradients, and the low vertical resolution of the sampling made the detection of inversion layers difficult. The cumbersome conventional systems are gradually being replaced by minisonde systems [1]. Minisondes are light and easy to launch, and the ground stations are portable and can be deployed virtually anywhere, so that horizontal resolution can be greatly improved. Similar improvements in tethered equipment have made boundary layer measurements easier and more reliable.

However, the full potential of modern minisonde systems cannot be exploited without a suitable method for determining the effect of the refractivity structure on electromagnetic propagation. Mode theory methods cannot cope with experimental data, and have to rely on template matching to idealized piecewise linear profiles [2] when an elevated layer is present. This means that the heights of top and bottom of the layer have to be calculated from the measured metorological profiles. Unfortunately, because of the imperfections of the sensors, the prediction of the location of the top and bottom of ducts becomes totally unreliable for typical sensor accuracies [3], so that mode theory methods are useless for real-time predictions. The Parabolic Equation Method does not have these drawbacks, and can be used to compute field-strength in a rectangular domain, from essentially arbitrary refractivity data [4], [5]. This opens up new possibilities for forecasting microwave propagation in the troposphere. The experimental facilities available at Rutherford Appleton Laboratory for meteorological measurements have been used to provide refractivity profiles in various propagation conditions. Field-strength forecasts are then calculated with PCPEM, a desk-top implementation of the Parabolic Equation Method, and in some cases the predictions can be tested against measured RF signals.

2. EXPERIMENTAL DATA

An experimental campaign to measure the refractive index structure at various sites along the U.K. coast was started in September 1988. The purpose of the campaign is to build up a catalogue of

meteorological events leading to anomalous propagation, and to characterize the refractivity structure in the boundary layer. Portsmouth was chosen for the first part of the campaign because it is the location of the receivers of several transhorizon links across the English Channel, which are part of project COST 210 on "The influence of the atmosphere on interference between radiocommunication systems at frequencies above 1 GHz". The objectives of COST 210 have been presented in [6], and a detailed description of the COST 210 transhorizon network across the English Channel can be found in [7]. The map in Figure 1 shows the location of the links. In order to study the effects of height diversity, four links have been set up at 11.6 GHz, using two transmitters located at Cap d'Antifer on the French coast and two receivers on the English coast. The layout of the experiment, with the high-high, high-low, low-high and low-low links, is shown in Figure 2. Signal strength is digitized and recorded at 1 Hz on all four links. A remote interrogation system is used to monitor the links. If an anomalous propagation event is detected, meteorological profiles can be measured using ADAS sondes.

The Atmospheric Data Acquisition System has been developed by Atmospheric Instrumentation Research inc. (A.I.R.) for the Naval Air Development Center (NADC) in the United States. The ADAS receiver is a portable unit, weighing 5.3 kg, which can be either mains powered or run on an internal battery giving it an autonomy of 4 hours. It receives and processes data measured by free sondes or tethered sondes. The free sonde (airsonde AS-3A) consists of a bead thermistor for temperature measurements, an aneroid cell for pressure measurements and a carbon hygristor for relative humidity measurements. It weighs 150 g. Data from all sensors are transmitted every 5 seconds, giving a 12.5 metres vertical resolution with a 100 g balloon ascending at the rate of 2.5 metres per second. The tethered sonde (tethersonde TS-3A-SP) measures dry and wet bulb temperatures with two matched thermistors, pressure with an aneroid cell, wind speed with a three-cup anemometer and a light-chopper tachometer, and wind direction with a magnetic compass. It weighs 225 g, and transmits data every 10 seconds. It can be used with a small tethered balloon ($2.25\,m^3$) which can be deployed fairly easily for boundary layer measurements.

Figure 3 shows the synoptic chart for 8 September 1988 at 1800 GMT. On that day, the English Channel was at the junction between two anticyclonic air masses located over Scandinavia and the Atlantic Ocean. Ground pressures in the Channel area were average, and temperatures on the South Coast were very warm despite the presence of a cold front. Conditions were ideal for anticyclonic advection of warm dry overland air from the continent over the sea. The transhorizon links were quite active, and a minisonde was launched at 1500 GMT. Figure 4 shows the measured refractivity profiles. A very low inversion layer was present, and the beginning of a structure which is probably the evaporation duct was measured at the start of the ascent. As the lowest point of the sonde data was at 12 m, the structure in the first 12 metres had to be extrapolated and a standard evaporation duct profile was assumed. A tethered sonde is now used to characterize the meteorological structure of the boundary layer. The recorded signal strength for that day is shown in Figure 5 for the four links, as transmission loss relative to free space in decibels against time in hours. The relatively narrow-band fluctuations and the high signal level are typical of a ducting event. The measured refractivity profile and the system parameters for the two transmitters were then used as input data for two runs of PCPEM [5] in order to generate two-dimensional field strength results, and transmission loss relative to free space was extracted along the four paths of interest. Even though no smoothing was applied to the rather noisy refractivity data, there were no numerical difficulties with PCPEM, which emphasizes the robustness of the method. Figure 6 shows transmission loss relative to free space (dB) against range (km) for the four links. The calculated values at a range of 152 km (low receivers) and 159 km (high receivers) are in excellent agreement with the data recorded at 1500 GMT, even though horizontal homogeneity had to be assumed for lack of meteorological data.

On 31 January 1989, measurements of the boundary layer were carried out with the ADAS tethersonde, on the beach at Portsmouth. Figure 7 shows the dry and wet bulb temperatures measured during one of the descents. The scatter is mostly due to horizontal inhomogeneities, since the 16 knots wind caused occasional horizontal movements of the tethered balloon. Smooth curves, also shown in Figure 7, were fitted to the data points. The resulting M and N profiles are shown in Figure 8. They are typical of an evaporation duct structure, although only crudely reminiscent of the classic log-linear shape obtained from Monin-Obukhov similarity theory [8]. Since the atmosphere is dispersive in the

higher part of the microwave spectrum and at millimetre wavelengths, the refractivity values should in fact be calculated for each frequency of interest, to take into account the dispersion and absorption due to oxygen and water vapour lines. Liebe's model [9] can be used to calculate the complex refractivity at any frequency up to 300 GHz from the meteorological profiles, and in turn this is used as input for PCPEM. Hence differential effects of the vertical water vapour profile can be studied. Because the temperature during the measurements was relatively low (7.4° C at sea level), the saturated water vapour density was only 8.2 g/m^3 at sea level and the effects are not as noticeable as at higher temperatures. Figures 9 and 10 shows 2-dimensional field strength diagrams at 18 GHz and 100 GHz respectively, for antennas located at a height of 10 m. The contours are in dB relative to a fixed offset at zero range, which depends on the power into the antenna. The plots are drawn in 4/3 Earth coordinates, so that rays would be straight lines in a standard atmosphere. At 18 GHz, the duct causes strong trapping of the lobes located just above it, and a complex pattern in the first 10 metres or so above the sea. Absorption effects are present because of the water vapour line at 22 GHz, but only cause a drop of 1.3 dB at most in the field at low altitudes, where the water vapour density is highest (at sea-level, the attenuation has a value of 0.065 dB/km). At 100 GHz, absorption is higher (0.55 dB/km at sea level) and the trapping is accompanied by shrinking of the contours, more marked at the lower altitudes.

3. MEASUREMENT ERRORS AND FIELD-STRENGTH PREDICTIONS

Measurement errors inevitably lead to fluctuations in field-strength predictions. Given the fact that PCPEM copes well with noisy data, the quality of the predictions can be assessed by numerical simulations. Here we assume a deterministic model for the medium, with a random element caused only by noise in the measurements. A more complex treatment is needed to model variability of the medium itself, for example to study scintillation effects. The statistical processes chosen for the simulations depend on the measuring instruments. Here we will concentrate on the case where the sensors measure pressure, temperature and relative humidity, but the method is easily adaptable to the case where other physical quantities are measured (wet bulb temperature for a tethered sonde, resonant frequency for a refractometer). In the current generation of minisondes, most systematic errors, including lag errors and hysteresis effects, can be calibrated out satisfactorily at the pressure levels of interest for refractivity measurements. Vertical resolution is now adequate, thanks to the lower rise rate of the small balloons which are used with the minisondes, so that sampling errors are virtually eliminated. However, random errors caused by the lack of accuracy of the sensors cannot be avoided. Here the random errors in the pressure, temperature and relative humidity measurements are modelled as independent Gaussian variables, and the measurements at different heights are also assumed to be independent. The standard deviation for each physical parameter is the accuracy specified by the manufacturer of the system under consideration. Once the ground-truth profiles for the meteorological variables have been chosen, a random number generator is used to obtain independent Gaussian variables of mean zero with the required standard deviations. These are added to the ground-truth profiles to obtain noisy profiles of pressure P, temperature T and relative humidity h. As height is calculated as a function of pressure, this process also introduces noise in the height profile. At low pressure levels, this method of calculating height introduces quite large errors in the profiles and a tracking radar should used [10], but this is not a problem at the pressure levels of interest.

The saturated water vapour pressure e_s is calculated using Tetens' approximation to the Clausius-Clapeyron equation

$$e_s = 6.105 \exp\left(25.22 \times (\frac{T - 273.2}{T}) - 5.31 \ln\left(\frac{T}{273.2}\right)\right), \tag{1}$$

where e_s is in millibars and T in Kelvin; refractivity N is then given by

$$N = 77.6\frac{P}{T} + 3.735 \times 10^5 \frac{e}{T^2}, \tag{2}$$

where $e = h \times e_s$ is the partial water vapour pressure. In this manner, a arbitrary number of noisy refractivity profiles can be generated from a set of ground-truth meteorological profiles. Some of the simulations were carried out using the accuracy parameters of the ADAS airsonde, others with the parameters of the VAISALA minisonde which is now used in several countries for routine synoptic soundings. Table 1 gives the accuracies used for the simulations.

Table 1: Accuracies of radiosonde sensors

	Pressure	Temperature	Humidity
AIR AS-3A	3 mb	0.5° K	5%
VAISALA RS 80	0.5 mb	0.2° K	2%

Once the set of noisy profiles has been generated, PCPEM is run with each of them to obtain a set of noisy field-strength predictions. To get a flavour of the type of conclusions that can be drawn from these statistical simulations, two examples are presented here. In each case, the frequency is 3 GHz. The first example uses ADAS measurements. Figure 11 shows a synoptic chart for 31 January 1989 at 1800 GMT. An exceptionally stable anticyclone was centred over France and Germany, with very high pressures in the English Channel and mild temperatures for the season. However the anticyclone was beginning to move East, and pressures were dropping steadily throughout that day. Figure 12 shows the temperature and relative humidity profiles measured by the ADAS airsonde at 1530 GMT, and Figure 13 shows the corresponding N and M profiles. The inversion layer, probably caused by subsidence, is about 100 m thick, between altitudes of 630 m and 730 m above mean sea level. This is too high to affect ground-based transmitters very much, and indeed there was little activity on the COST 210 links at that time. However, an intense anomalous propagation event, which lasted a few hours, started within an hour of the ascent, presumably because the subsidence layer had descended to lower altitudes after sunset. For the statistical simulations, the antenna was placed at a height of 700 m in order to maximize anomalous propagation effects. The field-strength predicted from the ADAS measurements, taken as ground-truth, is shown in Figure 14. The duct causes the separation of the beam into two distinct parts, with quite a complex structure in the layer. Figure 15 shows an example of a noisy run. Although the fine structure of the field is different, the main features are preserved. The assumption of independence of the errors at different heights generates very irregular profiles, hence the ragged aspect of the contours. By comparison, Figure 14 shows quite smooth contours, although it is itself based on unsmoothed experimental data. In fact the assumption of independence, which is used to simplify the study, leads to rather large error estimates for the predictions.

Because the field is in fact a realization of a 2-dimensional random process, it is quite difficult to present the statistical results in a conventional manner. One way to form a judgment on the usefulness of the predictions is to fix an acceptable error level, say 6 dB, and to calculate the probability that the prediction is in error by more than this level, at each range and height of interest [11]. The reference here is not the mean of the statistical runs, but the ground-truth field. Since the field-strength is unlikely to follow a textbook probability law, all the probabilities were estimated directly from the sample formed by the noisy runs. Figure 16 shows the probability that the prediction is in error by more than 6 dB, estimated from 100 runs. The darker areas, which correspond to a higher probability of error, are located in the hole just above the duct, and in the region where the field has a complex structure, in the duct or just below it at long ranges. The exact location of sharp localized features of the field is very sensitive to noise in the refractivity data, hence these regions of low confidence in the predictions. Zones where the probability of a 6 dB error is more than 50% are very limited in extent, and the probability of a 6 dB error is less than 20% in the greater part of the frame, which indicates fairly good quality of the predictions. Figure 17 shows histograms for the field at a height of 600 m and ranges of 40, 60, 80 and 100 km. The spread of the distributions is at most 16 dB, which is quite encouraging since these points lie in one of the more unstable regions. The distributions are asymmetric, and vary considerably with range, showing the difficulty of trying to represent the field as a classic random process. In particular, a lot of information would be lost if the field was only represented by its mean and standard deviation. On each histogram, the ground truth value is indicated by a cross, and the mean value by a circle. Clearly, the mean field is a biased estimator of the ground-truth field. This again reflects the fact that sharp features of the field change their location in each statistical run, and tend to get smoothed out when the mean is taken.

The second example uses artificial meteorological profiles from [3], which have been chosen to be representative of situations where two types of ducts are present. Figure 18 shows the temperature and humidity profiles, and Figure 19 shows the resulting N and M profiles. A vertical resolution of 20 m was chosen, and a hundred noisy profiles were generated in the manner described above, using the VAISALA RS80 system specifications. The 3 GHz antenna was placed at a height of 100 m. The ground-truth field

is shown in Figure 20. The surface-based duct causes strong trapping, and the elevated duct causes distortion of the higher part of the beam. Figure 21 shows the mean field. The finer structure inside the surface duct has disappeared, and the location and extent of the lobes above the surface duct is slightly different from ground truth, but the main features are preserved. Figure 22 shows a probability map for an error threshold of 6 dB. Again, the probability of error is greater in areas where the structure is more complex, in the duct and in the leakage area just above it. As in the first example, the probability of a 6 dB error is less than 20% except in quite limited areas. Figure 23 shows a probability map for an error threshold of 3 dB. Of course the situation is now much worse, with quite a large area where the probability of a 3 dB error is more than 20%, but this threshold is perhaps unrealistic in view of the other sources of error (lack of information on the horizontal refractivity structure, surface effects). Although these two cases are fairly representative of strong ducting conditions, further numerical tests should be conducted with other meteorological data sets in order to assess sensitivity to measurement errors in a variety of situations.

4. CONCLUSIONS

The Parabolic Equation Method is sufficiently robust to cope with noisy refractivity input data, in the microwave as well as the millimetre wave region of the spectrum, and the predictions are in good agreement with measured RF signals. Because of random fluctuations in the measurements and in the medium, the electromagnetic field cannot be described in a deterministic way. These stochastic effects can now be quantified, in order to assess their impact on propagation forecasts. Statistical simulations indicate that field-strength predictions are relatively insensitive to radiosonde errors. The main features of the field are essentially unchanged by noise in the refractivity data, but the fine structure cannot be predicted reliably. Whenever possible, the forecasts should be accompanied by a probability map indicating levels of confidence in the predictions.

5. ACKNOWLEDGMENTS

This work has been supported in part by the Procurement Executive, Ministry of Defence. The ADAS receiver used for the meteorological measurements has been loaned to RAL by the Air Force Geophysical Laboratory (AFGL).We wish to thank our RCRU colleagues for their support, and in particular C.J. Gibbins for his computer implementation of Liebe's model.

6. REFERENCES

[1] Nash, J. and Schmidlin, F.J., WMO international radiosonde comparison, Final Report, WMO Instruments and Observing Methods, Report No 30, **WMO/TD-No. 195** (1987).

[2] Baumgartner, G.B. , Hitney, H.V. and Pappert, R.A. , Duct propagation modelling for the integrated-refractive-effects prediction system (IREPS), *IEE Proc. F* **130** (1983), 630-642.

[3] Gathman, S.G., Sensitivity of Refractive Index Profile Models to Relative Humidity Sensor Errors, Naval Research Laboratory Memorandum Report 5758 (1986).

[4] Craig, K.H., *Elec. Lett.*, **24** (1988), 1136–1139.

[5] Craig, K.H. and Levy, M.F., this publication (1989).

[6] COST 210 Management Committee, *IEE Conf. Pub.*, **274** (1987), 351-354.

[7] Vilar, E. , Spillard, C., Rooryck, M., Juy, M., Barber, P.C. and Hall, M.P.M. , *Elec. Lett.*, **24** (1988), 1205-1206.

[8] Rotheram, S. , Microwave duct propagation over the sea, *IEE Conf. Pub.* **219** (1983), 9-13.

[9] Liebe, H.J. , An updated model for millimeter wave propagation in moist air, *Radio Science*, **20** (1985), 1069-1089.

[10] Schmidlin, F.J., Can the standard radiosonde system meet special atmospheric research needs ?, *Geophys. Res. Lett.*, **9** (1982), 1109-1112.

[11] Levy, M.F. and Craig, K.H., Case studies of transhorizon propagation: reliability of predictions using radiosonde data, IEE Conf. Pub. **301** (1989).

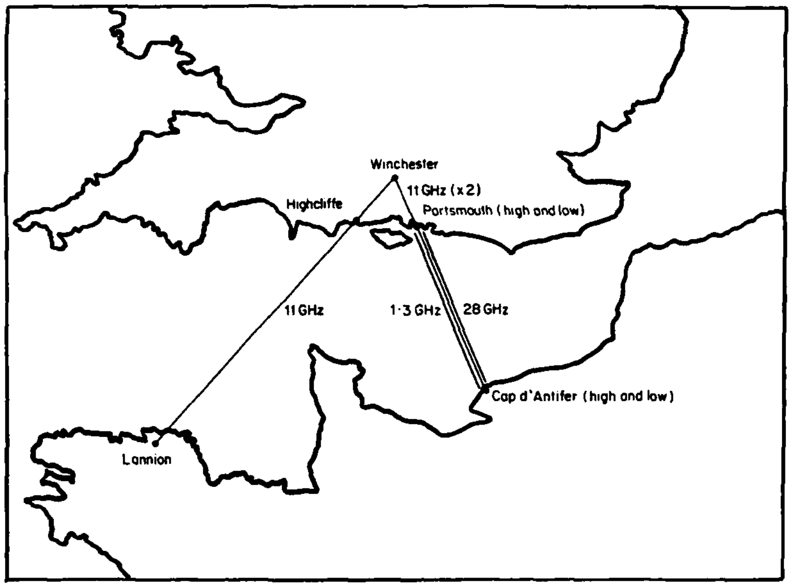


Figure 1. COST 210 transhorizon links across the English Channel.

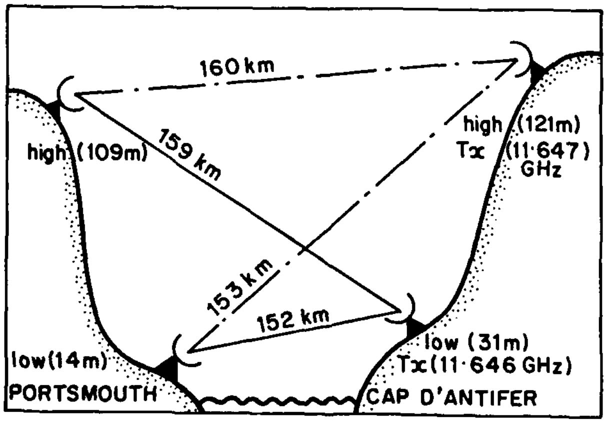


Figure 2. Height diversity experiment.

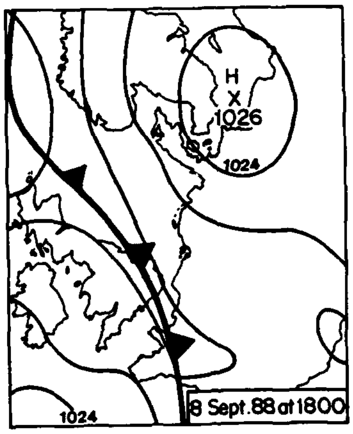


Figure 3. Synoptic chart, 8 September 1988.

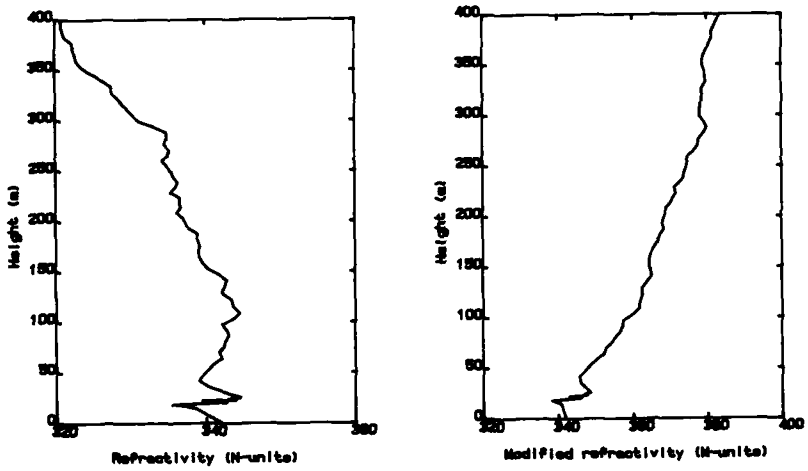


Figure 4. Minisonde refractivity profiles, 8 September 1988, 1500 GMT.

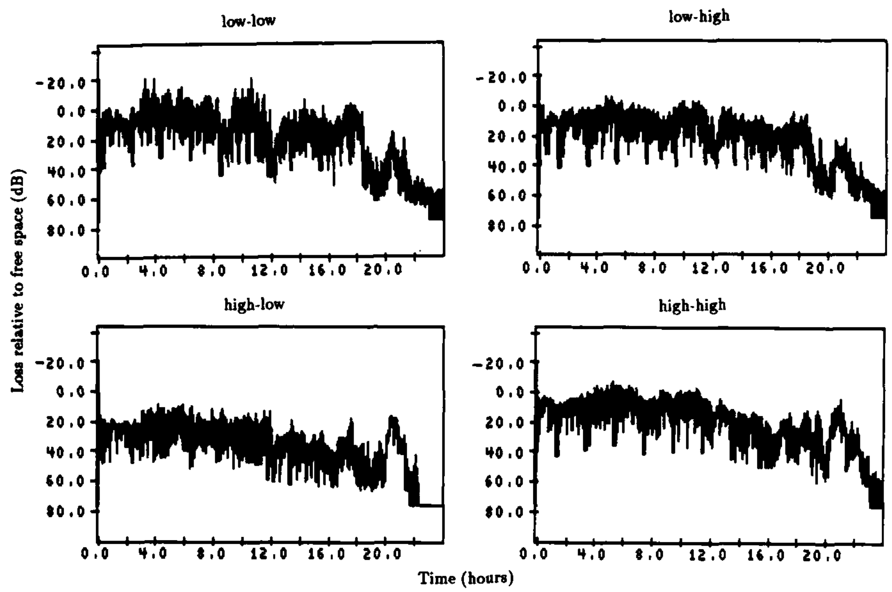


Figure 5. Signal recorded on COST 210 links, 8 September 1988.

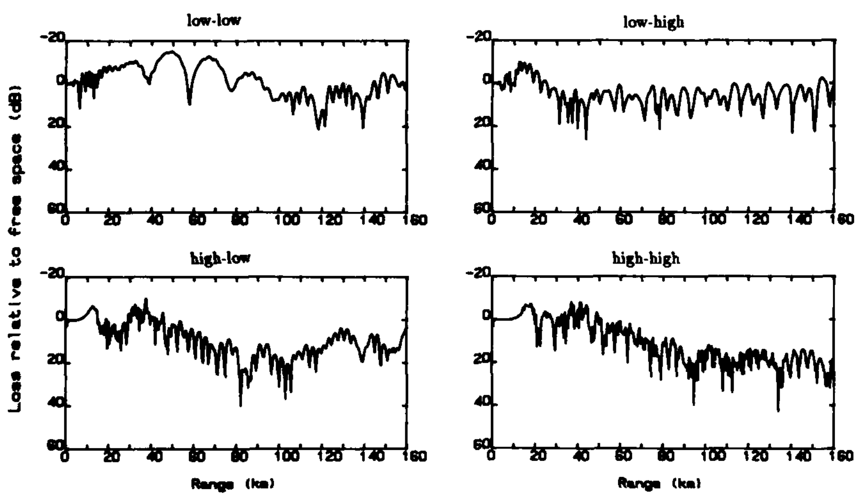


Figure 6. Signal calculated from minisonde data at 1500 GMT.

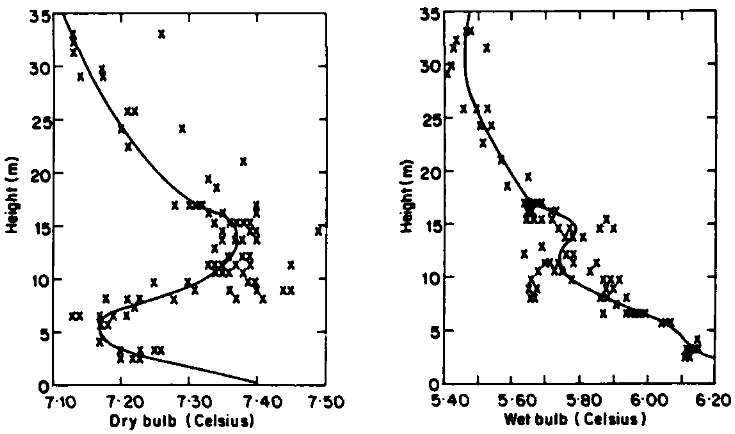


Figure 7. Tethersonde profiles, 31 January 1988.

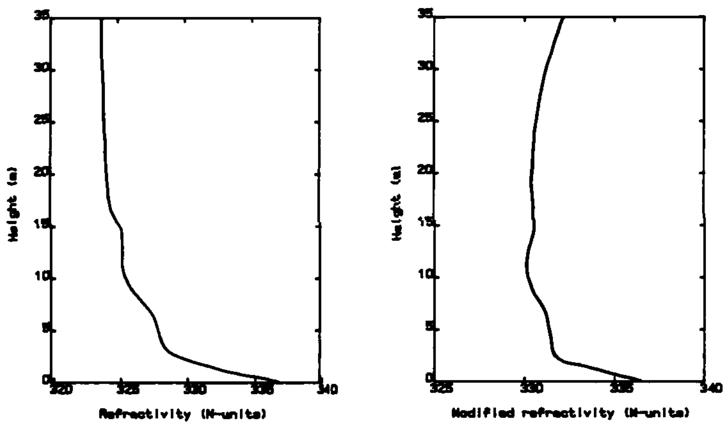


Figure 8. Refractivity profiles from smoothed tethersonde data.

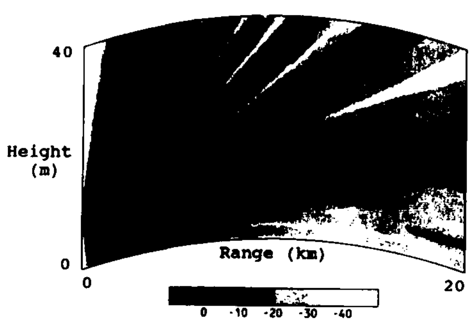


Figure 9. Field of an 18 GHz antenna at 10 m, tethersonde data.

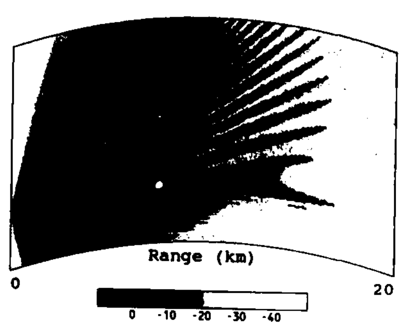
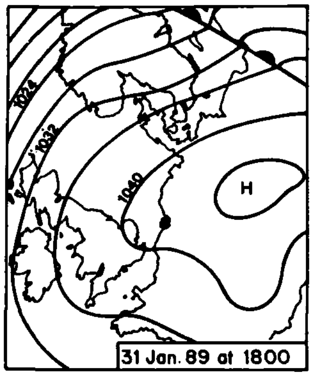


Figure 10. Field of a 100 GHz antenna at 10 m, tethersonde data.



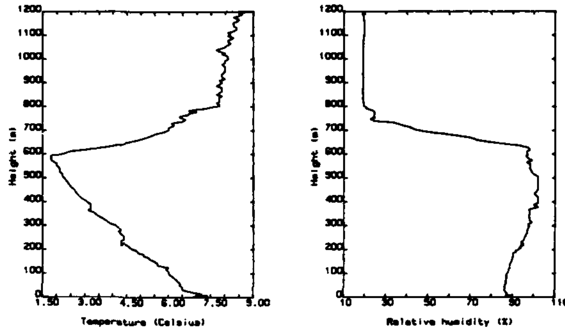


Figure 11. Synoptic chart, 31 January 1989. Figure 12. Minisonde profiles, 31 January 1989, 1530 GMT.

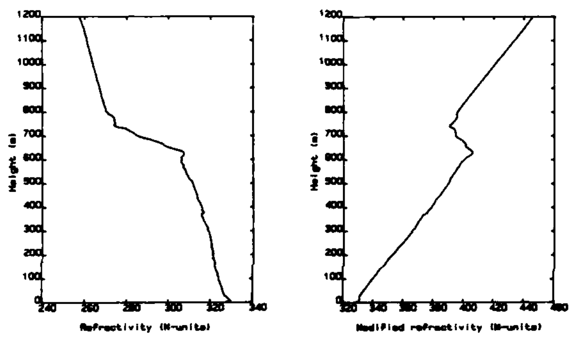


Figure 13. Refractivity profiles, 31 January 1989, minisonde data.

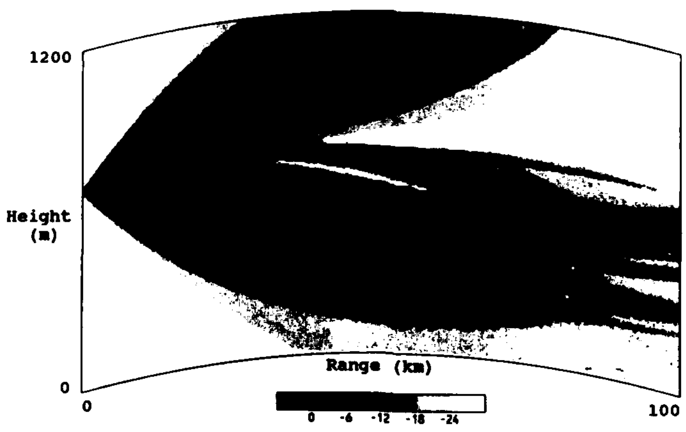


Figure 14. Field of a 3 GHz antenna at 700 m height, ground truth from minisonde data

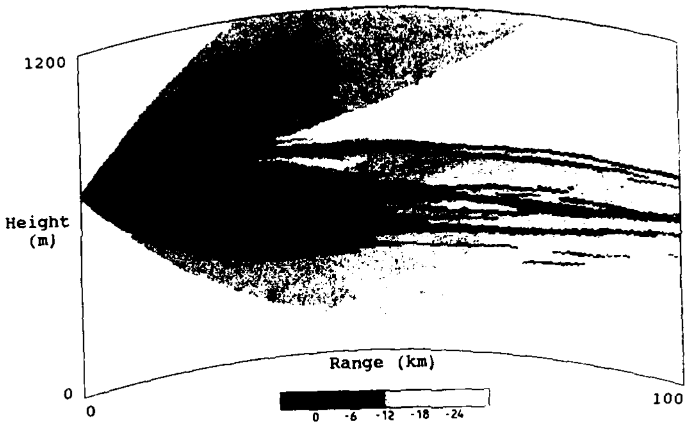


Figure 15. Noisy field from one of the statistical runs.

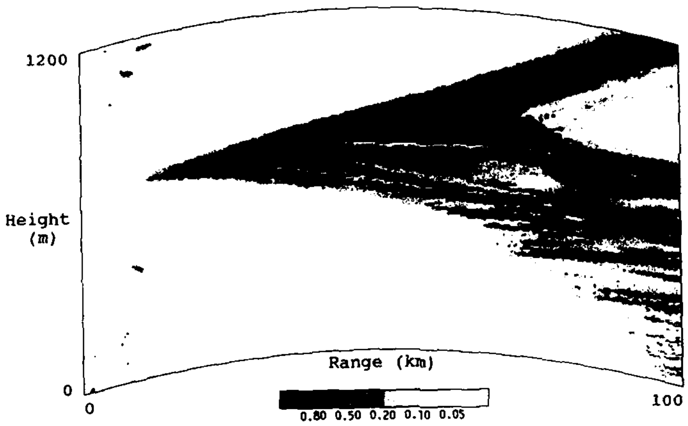


Figure 16. Probability of error greater than 6 dB, 100 statistical runs.

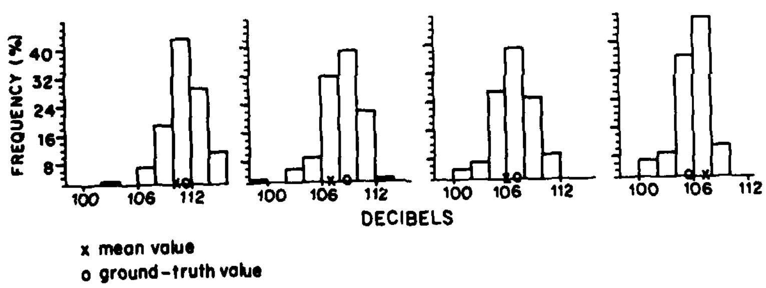


Figure 17. Histograms for field at 600 m height and 40, 60, 80, 100 km range.

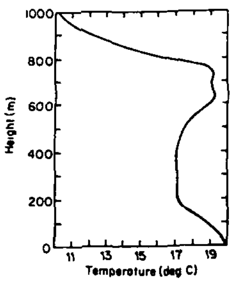


Figure 18. Meteorological profiles, second statistical simulation.

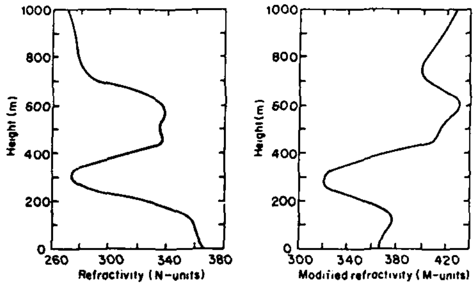


Figure 19. Refractivity profiles, second statistical simulation.

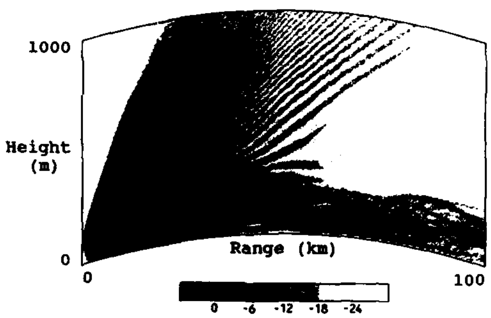


Figure 20. Field of a 3 GHz antenna at 100 m height, ground truth from Figure 18 data.

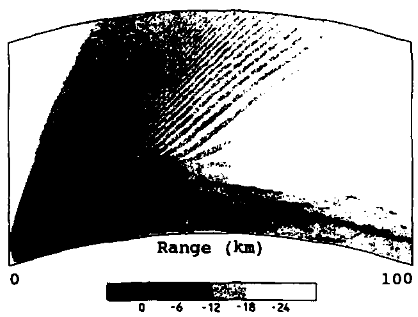


Figure 21. Mean field, 100 statistical runs.

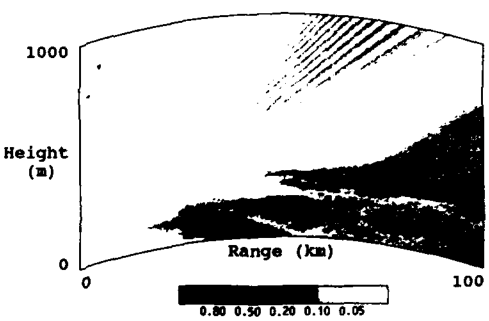


Figure 22. Probability of error greater than 6 dB, 100 statistical runs.

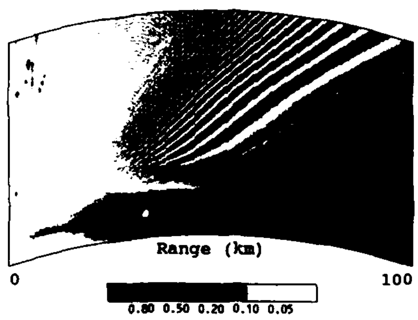


Figure 23. Probability of error greater than 3 dB, 100 statistical runs.

FORWARD SCATTER PROPAGATION PATH LOSS TESTING USING SURROGATE TERRAIN IN THE 100 to 1500 MHz REGION

by

Louis L.Taylor
The Mitre Corporation
145 Wyckoff Road
Eatontown, New Jersey 07724
USA

ABSTRACT

Numerous path loss calculation models have been developed for links employing forward scatter propagation. Variances in propagation loss prediction between these models and the loss experienced when a link is actually established arise because of the manner in which the models represent the large variety of absorbers and reflectors present in the actual terrain. A method to enrich a model with actual measurements that produces a more accurate method of predicting propagation loss is described. The method is based upon selecting a surrogate section of terrain for the actual terrain over which a propagation loss determination is desired, then measuring the forward scatter loss variation through this surrogate terrain. The loss value is then used as a factor in the simplified mathematical model used for the desired prediction.

Examples are given of the use of this technique to assist in the design and implementation of a wide area VHF/UHF radio network serving a large fleet of mobile units.

Through use of this method, one learns to judge when existing models may be used and when measurements are indicated, how to estimate the effects of different types of terrain and foliage upon propagation, and how these effects change in different seasons of the year.

INTRODUCTION:

Ever since Bullington presented his monumental work (1), those interested in propagation have been refining the process of predicting the propagation loss of a path that includes forward scatter phenomena. The works of Norton (2) and then Rice, Longley et al. (3) were great steps in the process; their work was enhanced by Hufford (4), Egli (5), and Okumura (6), to name a few. Computer models based on the work of these, and other individuals are proliferating.

All of these models have the problem that estimates of terrain factors (such as roughness, foliage, and scattering effects) must be reduced to numbers, by judgemental factors, and plugged into the models. The results of the model, with these judgemental factors included, are accepted as gospel.

Difficulty with such a procedure arises from the fact that the accuracy of the predictions is dependent on the wisdom of the estimator and no tool has yet been developed to determine accurately the effects of a grove of trees, a group of buildings, or a series of low hills upon the propagation attenuation of a particular path. Even if such a tool were available, it would be a time-varying device since the received signal between two fixed sites communicating over a forward scatter path produces a signal that has considerable variation in magnitude over any significant period of time.

This paper provides you with some tools to augment rationale and simplify the process of predicting the propagation loss of a forward scatter path.

LIMITATIONS:

There are propagation paths, and then there are propagation paths. They vary in characteristics from simple line-of-sight paths, with no reflections (which are experienced only in free-space); through paths that use ionospheric or tropospheric reflection; to paths through the earth and paths that employ railroad tracks to guide the electromagnetic energy. Many papers on the subject of propagation try to cover several, or more, different types of propagation paths and a very broad spectrum of frequencies. As a result, the engineer who specializes in fields that normally do not include electromagnetic propagation, but who needs to learn something about a particular mode of propagation is dissuaded from further investigation by the complexity of the treatment of the subject.

This paper describes one particular mode of propagation. This is the propagation phenomenon that is experienced in the frequency range from 100 to 1500 MHz between two sites on the earth's surface that are at a distance apart which precludes the direct line-of-sight transmission illustrated in figure 1a, but enables communication by means of forward scatter from reflecting (scattering) objects located between the two sites. This mode of propagation is illustrated in figure 1b.

It will be assumed that the two sites have antenna heights anywhere from shoulder height to 100 meters, but these limitations are not absolute. It is absolute that the antennas be above ground but not located in aircraft.

BACKGROUND:

Consider a propagation path between a transmitter and a receiver composed of two free-space paths and a forward scatter path, as diagramed in figures 1b and 2. The three separate segments of the propagation path are denoted as D_1, D_2, and D_3 in figure 2. Distance D_1 is measured from one antenna to the point where the radiation of the antenna is tangent to the surface of the earth. Distance D_2 is the same for the other antenna, and D_3 is the distance between the two tangent points. The discussion in this paper holds true when D_3 is zero as illustrated in figure 3a, or even negative, as illustrated in figure 3b, to the limit where it is so negative that 0.6 of the first Fresnel zone clearance (0.6F) is achieved between the direct path between the two antennas and the surface of the earth, as illustrated in figure 1a. When this limit is exceeded, the propagation path loss is considered to be free-space loss, and the value in dB is given by the formula:

$$L_{los} = 32.47 + 20\text{Log}_{10}(f \times D), \qquad (1a)$$

where f is frequency in MHz, and D is the total path length in km. If it is desired to express the path length in statute miles, the expression becomes:

$$L_{los} = 36.6 + 20\text{Log}_{10}(f \times D), \qquad (1b)$$

or using natural logarithms: $L_{los} = 36.6 + 8.6859\text{Ln}(f \times D)$,

If the heights of the two antennas are equal, or nearly so, over a smooth earth the value 0.6F in the middle of the path can be obtained from the formula:

$$0.6F = 164.3\ \text{SQR}(D/f), \qquad (2a)$$

where D and f are the same as in (1a) and the value of 0.6F is in meters. If it is desired to use statute miles instead of kilometers, this becomes:

$$0.6F = 684\ \text{SQR}(D/f). \qquad (2b)$$

For those who have not met Fresnel, the first Fresnel zone is an ellipsoid, with the two antennas as foci. The surface of this ellipsoid is the loci of all reflecting points where the energy propagated from one antenna to the other via the reflecting point will be 180 degrees out of phase with the energy propagated along the direct path between the antennas. The formula for this ellipsoid is:

$$F = 547.8\ \text{SQR}\ [(d_1 \times d_2)/(f \times D)], \qquad (3a)$$

where d_1 and d_2 are the distances from each antenna to a point on the axis of the ellipsoid, F is one-half of the diameter (i.e., radius) of the ellipsoid at that point. The distance between the two antennas is D, i.e., $D = d_1 + d_2$. The dimensions for D, d_1, and d_2 are kilometers, f is MHz, and F is meters. If one wishes to use feet and statute miles, the expression becomes:

$$F = 2280\ \text{SQR}\ [(d_1 \times d_2)/(f \times D)]. \qquad (3b)$$

Formula (3) is general, but formula (2) is specific and can provide a quick estimate of the type of propagation present (free-space or forward scatter) for a smooth earth if the antenna heights at the two ends of the path are similar. If the antenna heights are very dissimilar, the value of 0.6F should be calculated through use of formula (3a) or (3b). Also, if smooth earth conditions do not exist, visual sighting or map methods must be employed to determine if free-space conditions prevail. The point is that a distance of 0.6F must exist between the direct line of sight between the two antennas and the closest approach of this line to the earth's surface for the propagation to be considered as free-space.

For all of these propagation path calculations, the surface of the earth is considered to be 4/3 of the actual earth radius of 5,333 statute miles, or 8,583 km to compensate for the average bending effect of the atmosphere on electromagnetic radiation.

MODELS

Bullington's empirical nomographs were extremely useful before the advent of computers, and they can still be used today by someone who does not have ready access to a computer and the appropriate programs. Reference 1 is available in many technical libraries, but the nomographs have been reproduced in many current publications (references 7 and 8 are two examples). Many computer models are available, most of which are based upon the pioneering work of Norton, Bullington, Longley, Rice, and Egli. Most of these computer models require one to input, in addition to the known variables, values that represent terrain roughness and foliage; some even require an estimate of other reflecting objects such as buildings and objects to either side of the propagation path between the two ends of the propagation link. This information must be estimated from topographic maps, photographs, and/or personal knowledge.

For those of you who do not have ready access to Bullington's nomographs nor a computer model, but who do have an IBM-compatible computer, you might be able to make use of the simple model on the diskette that was prepared in conjunction with this paper. Copies are available here at the podium after this series of papers, or can be made from the formulas presented in this paper, or I will send copies to you if you will send me a self-addressed, stamped diskette mailer with a diskette in it formatted on your machine.

This model is based upon Bullington's nomographs in the region from 30 to 1500 MHz. It was derived by converting Bullington's nomographs to a set of two dimensional curves then generating a simple set of equations to fit the curves. In spite of the title of this paper, the following equations track Bullington's nomographs to within a couple of dB well below 100 MHz, down to about 30 MHz. They were extended to this frequency to provide coverage for the lower VHF frequencies.

$$L_1 = X/(d_1)^2 + Y*(d_1)^2 + 18.25 \qquad (4)$$

$$X = -1511212/(f)^2 + 70669/(f) - 26.65 \text{ and}$$
$$Y = 3.33E-10*(f)^2 + 1.46E-06*(f) + 5.58E-05$$

$$L_2 = -G/(d_2)^2 + H/d_2 \qquad (5)$$

$$G = -15.69*[Ln(f)]^2 + 239.7*Ln(f) - 913.8$$
$$H = -47.21*Ln(f) + 375.3$$

$$L_3 = M*(d_3)^2 + N*d_3 + O \qquad (6)$$

$$M = -.00171*[Ln(f)]^3 + .0272*[Ln(f)]^2 - .1379*Ln(f) + .2226$$

$$N = .1173*[Ln(f)]^3 - 1.7507*[Ln(f)]^2 + 8.6412*Ln(f) - 13.4204$$

$$O = -.3949175*[Ln(f)]^3 + 6.348961*[Ln(f)]^2 - 32.71855*Ln(f) + 53.1799,$$

where d_1, d_2, and d_3 are the path segments illustrated in figure 2 expressed in miles (km x 1.61), and L_1, L_2, and L_3 are the propagation losses, in dB, over each of these segments. An approximation for the values for d_1 and d_2 can be obtained from the heights of the respective antennas and the expression:

$$d = SQR\ (2 \times h), \qquad (7)$$

where h is the <u>effective</u> antenna height (9) in feet (meters x 3.28).

The total propagation loss for a path is:

$$L_T = L_1 + L_2 + L_3 + L_{los}\,. \qquad (8)$$

When the model on the diskette is used, it will provide calculated values of d_1 and d_2 based on a smooth earth and will ask for the user's concurrence. If terrain conditions indicate a different path length the revised figure may be entered into the model. The model will then provide a calculated value of L_3, i.e., loss in the forward scatter region. Here is where judgement and experience are useful. The presence of trees, buildings, and rough terrain will produce a greater L_3 than the calculated value. One who has gained experience in this area can enter an increased value of L_3 to include the effect of trees and terrain.

If surrogate terrain has been used to measure the total path loss for comparison with the value for the same path provided by a model (as described in the following section) increase the value of L_3 by the amount that the measured value exceeded the value of L_T provided by the model. Otherwise, re-enter the value provided by the model to obtain the preliminary estimate of total loss.

SURROGATE TERRAIN:

Figure 4 is a plan view showing the line-of-sight coverage of two antennas of dissimilar heights, and the forward scatter area between them. A little rationale and visualization will lead one to realize that absorbing objects (e.g., foliage) in the forward scatter area will increase L_3. It has been found that foliage effects at 30 MHz are minimal, at 450 MHz they become appreciable, and at 1500 MHz they are practically prohibitive. Further rationalization will lead to the awareness that reflecting objects in the forward scatter area, even to one side of the direct line between the two antennas, will

The key to the use of surrogate terrain for path prediction is to measure the propagation loss over a path that has a forward scatter area similar in characteristics (type and amount of foliage, terrain type, and forward scatter path length) to the path in question. Then, calculate the loss over the surrogate path with a Bullington nomograph, the Longley-Rice Model, the model on the diskette discussed above, etc. The difference between the total loss obtained from calculations and the measured value of the loss is the error in the model. Then, using the same model, calculate the loss for the path in question and add the correction value obtained over the surrogate path to L_3 in the model to fine tune the calculations for the particular path in question.

Incidentally, another excellent model that is just becoming available is the Terrain-Integrated Rough-Earth Model, otherwise known as the TIREM model (10) which is available from the National Technical Information Service. If you desire to use this model be certain that you obtain Change 4 as previous editions of the reference do not contain the TIREM model.

Several precautions, learned from experience, are warranted here:

1. Trees and other foliage in the vicinity of either antenna add additional attenuation if the antenna is not above the trees. This attenuation at 150 MHz for a thin stand of hardwood will be around 3 dB, for heavy pine trees 5 dB, and heavy swamp vegetation will increase the value to about 7 dB. At 450 MHz these values increase to 6, 10, and 25. They are still higher at 1500 MHz and minimal at 30 MHz.

2. Greater accuracy from most models, including the one on the diskette, will be obtained if the antenna height is measured from the average height of the heavier vegetation. In the absence of vegetation, antenna height should be measured from average terrain height (9).

3. Foliage in the forward scatter area decreases the scattering and increases L_3. The amount of this effect is the subject of ceaseless argument among propagation experts and is best handled by actual measurements, such as the surrogate terrain method discussed herein.

4. Smooth terrain undulations in the forward scatter region provide less scattering (greater L_3) than sharp, jagged undulations.

5. A good indication of the magnitude of forward scattering is indicated by signal variations over a few wavelengths of change of path length. The greater the variation, the greater the forward scatter gain is a general rule.

FIELD STRENGTH MEASUREMENTS

The role of field strength testing has been much maligned, but extremely useful information can be gained from field strength testing, especially when it is combined with map study, visual investigation, and aerial or satellite

photographs. The advent of high-resolution satellite photographs coupled with the highly accurate topographic maps available today enables one to select surrogate terrain for testing that represents the actual terrain with confidence (11). The following precautions are offered to assist in avoiding the pitfalls often encountered in field strength measurements:

1. Mount the test antenna in a manner that enables true omni directional coverage, or use a directional antenna with known gain, and mount it so the design gain is unimpaired.

2. Calibrate, calibrate, calibrate! Calibrate all test equipment before and after the testing, and if possible, several times during the testing.

3. At the beginning of a test, vary range over a significant distance and see if the change in loss due to change in range is similar to the change in loss predicted by the model for the same change in range. If not, ducting or skip due to atmospheric inversion (or ionospheric anomalies) is present, and the testing must be delayed until the abnormal conditions vanish and only forward scatter is present. A balloon sonde or drop sonde used to measure the vertical gradient (M profile) is an excellent tool to detect atmospheric anomalies, but such a luxury is not often available at the time of the testing.

4. Measurement data should be collected while moving the receiver, or transmitter, through at least 100 wavelengths change in range. Movement must be slow enough so the recording device can track maximum and minimum signals. The greater the variation the smaller the number of predominant scattering objects affecting the path. The value of received signal strength used to determine path loss should be the average of the varying signal.

5. Frequency difference between the surrogate path and the actual path should not exceed ten or fifteen percent.

APPLICATION:

Figure 5 shows the tower locations and heights for a statewide VHF/UHF communication system serving all law enforcement and state government vehicles in one state of the United States. The tower locations were somewhat constrained by the availability of property owned by the State and available for use as a tower site. Transmitter power output and antenna gains were based on commercially available equipment. Base station transmitter power output was chosen to be 3 dB higher than mobile transmitter power output, and base station receiver sensitivity was 3 dB better than mobile receiver sensitivity; therefore, the system talk-out range was equivalent to system talk-back range. The selection of tower heights at each location were chosen to provide a 90 percent coverage, i.e, a usable signal strength would be received by a mobile unit at 90 percent of the locations in the state 90 percent of the time.

The Bullington smooth earth model was used to calculate the necessary tower height for each site by calculating twelve radials from each tower site,

plus additional radials where foliage or terrain indicated potential problem areas. The forward scatter loss in the model was increased for terrain variance from smooth earth by use of contour maps and the for foliage by use of satellite photographs. The satellite photographs were used, also, to estimate the additional loss that would be incurred by mobiles operating in forested areas. The satellite photographs made it an easy matter to differentiate between the type and density of the foliage which varied from moss-covered swamp evergreens in the southern part of the state to sparse hardwood stands in the northern part of the state. A composite infra-red satellite photograph of the entire state is shown in figure 6. High resolution satellite photographs of smaller areas are now available from several sources. One, for example, is the Spot Image Corporation at 1897 Preston White Drive, Reston, Virginia, USA.

Existing transmitters in the state operating in the VHF and UHF regions were used as surrogate emitters to obtain improved forward scatter terrain and foliage constants. Radial propagation path measurements from fourteen such transmitters enabled the creation of a library of forward scatter constants for the entire state. Figure 7 shows part of a sample recording run with received signal strength on the ordinate and range from the transmitter on the abscissa. Note the fine structure and the coarse structure in the plot. The fine structure shows peaks at one-wavelength intervals and the coarse structure is dependent upon scattering objects and height of the test unit. The one predominant coarse structure peak is the result of traversing a railroad overpass which increased the height of the mobile unit significantly. Annotations on the plot are calibrating notations made by the operator.

Figure 8 compares the Bullington smooth earth model, expressed in equations (4) through (8), with and without compensation for trees and terrain, with measured values over a 30-mile radial at 150 MHz. Note that when close to the transmitter the high angle of incidence negates the effects of foliage and terrain, but at about 5-miles range the measured value follows the model very closely. As one might expect, at extreme ranges the propagation loss is extremely sensitive to local ground elevation, as indicated by the measured values in the dip and on a little rise at about 30-miles range.

As a result of using the available VHF and UHF transmitters in the state to make surrogate measurements it was not necessary to employ any test equipment other than the instrumented mobile unit to make field strength tests and enrich the model. The result was a system that met the coverage requirements without incurring extra costs due to expensive testing nor excessive tower heights to assure the desired coverage.

CONCLUSION

Propagation models are a luxury, but often cannot be used due to computer system unavailability, or due to the ability to obtain the proper inputs for the model. Path loss testing over a path similar to the terrain of interest can produce forward scatter loss information that can be utilized, through the use of a crude model, to obtain an error between the model's prediction and the actual path loss. This magnitude of error can then be used in the same model with the constants for the desired path to determine more

accurately the loss over the desired path, i.e., formula (8) can be restated and used to calculate the total propagation path loss for the terrain of interest as:

$$L_T = L_1 + L_2 + L_3 + L_{los} + L_e, \qquad (8a)$$

where L_e is the difference between L_T measured over the surrogate terrain and L_T calculated by the model for the same path over the surrogate terrain.

References:

1. K. Bullington, "Radio Propagation Fundamentals," The Bell System Technical Journal, Volume 36, Pages 593-626, 1957.

2. K. A. Norton, Transmission Loss in Radio Propagation, NBS Technical Note Number 12, June 1959.

3. P. L. Rice, A. G. Longley, K. A. Norton, and A. P. Barsis, Transmission Loss Predictions for Tropospheric Communication Circuits, NBS Technical Note 101, Volumes I and II (revised), January 1967.

4. G. A. Hufford, "Modifications to the use of the ITS Irregular Terrain Model in the Point-to-Point Mode," U.S. Department of Commerce, NTIA Memorandum, January 30, 1985

5. J. Egli, "Radio Propagation Above 40 MC Over Irregular Terrain," Proceedings of the IRE, Pages 1383-1391, October 1957.

6. Y. Okumura, E. Ohmori, T. Kawano, and K. Fukuda, "Field Strength and its Variability in VHF and UHF Land Mobile Service," Review of the Electrical Communication Laboratory, Volume 16, Page 825, September-October 1968.

7. W. C. Jakes, Jr., Mobile Microwave Communications, John Wiley & Sons, Inc., 1954, page 86.

8. P. N. Saveskie, Radio Propagation Handbook, TAB Books, Inc., 1980, Pages 210, 211.

9. Ibid., Appendix 29, pps 480-482.

10. M. Weissberger, R. Meidenbauer, H. Riggens, and S. Marcus, Radio Wave Propagation: A Handbook of Practical Techniques for Computing Basic Transmission Loss and Field Strength, ECAC-HDBK-82-049, Available from NTIS as NTIS-A122090.

11. L. L. Taylor, "Practical Field Strength Measurements," Proceedings of the 1977 Carnahan Conference, Lexington, Kentucky.

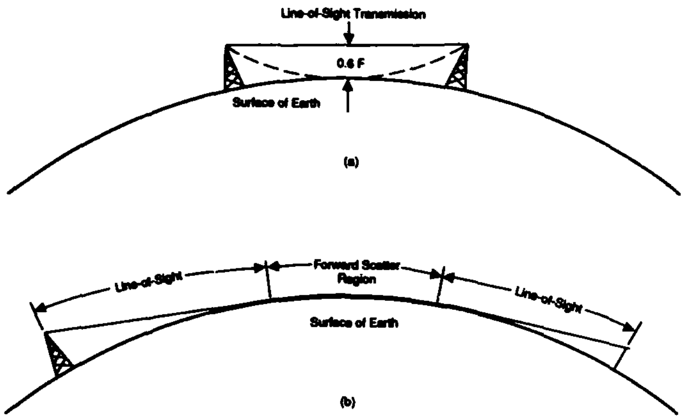


FIGURE 1
FORWARD SCATTER PROPAGATION

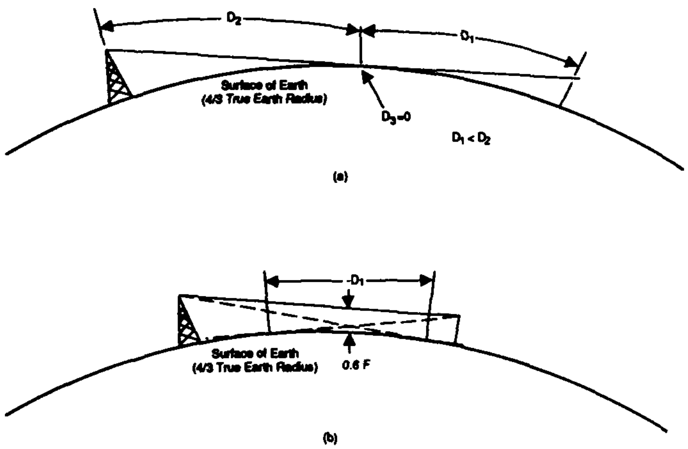


FIGURE 3
LIMIT OF FORWARD SCATTER PROPAGATION

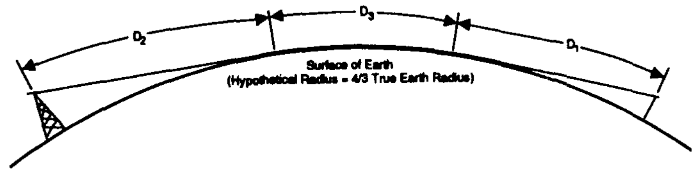


FIGURE 2
PATH SEGMENTS OF FORWARD SCATTER PROPAGATION

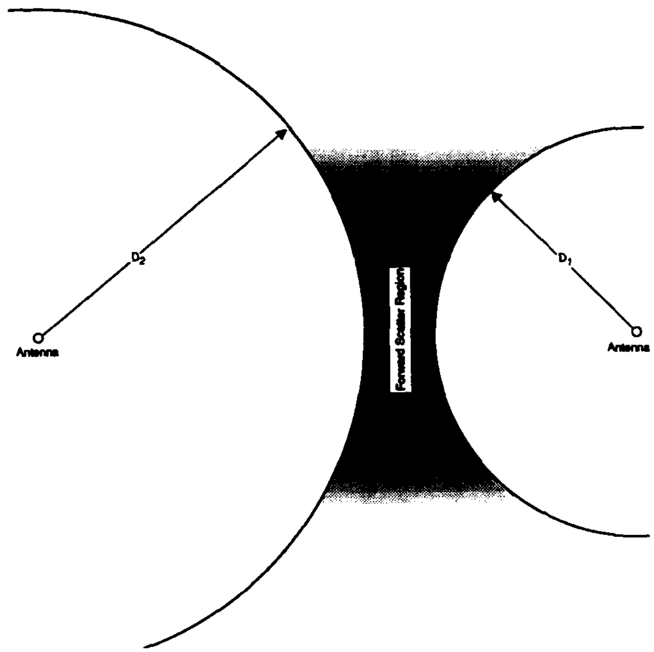


FIGURE 4
PLAN VIEW OF FORWARD SCATTER REGION

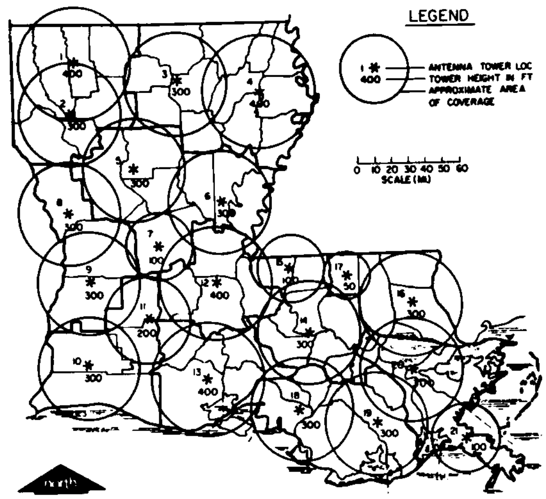


FIGURE 5
TOWER LOCATIONS AND HEIGHTS

FIGURE 6
INFRA-RED SATELLITE PHOTOGRAPH

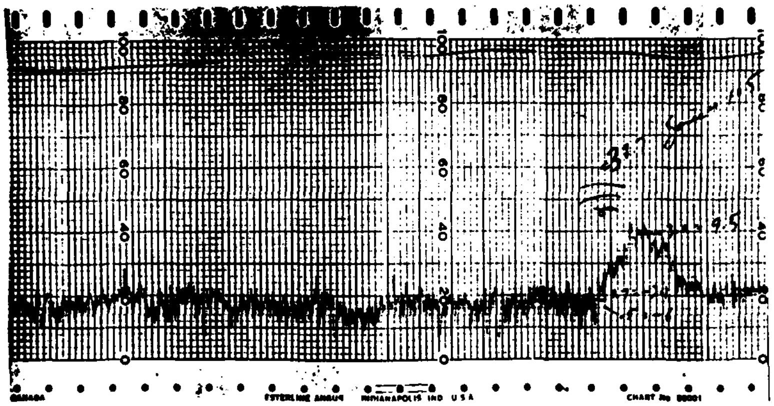


FIGURE 7
SAMPLE RECORDING

FIGURE 8
BULLINGTON VERSUS MEASURED LOSS

AMELIORATION DE LA QUALITE D'UNE LIAISON RADIOMOBILE A ETALEMENT DE SPECTRE

A. DANIEL
G. EL ZEIN, M. SALEHUDIN, J. CITERNE
Laboratoire Composants et Systèmes pour Télécommunications
UA au CNRS N° 834, INSA 35043 RENNES Cédex (France)

RESUME (NU)

Cette communication concerne l'étude expérimentale d'un dispositif permettant d'améliorer la qualité d'une liaison numérique radiomobile à étalement de spectre à l'aide de la diversité en prédétection par combinaison à gain maximal de multitrajets.

L'originalité du système présenté dans cette communication réside dans la réalisation d'un filtre transversal conçu à partir d'une ligne à retard multiprises à ondes acoustiques de surface (SAW) et adapté à la réponse impulsionnelle du canal, obtenue par un convoluteur (SAW). L'adaptation du récepteur aux paramètres du canal de transmission est réalisée avec des coefficients de pondération instantanés. Les multitrajets sont ensuite combinés afin d'éliminer au mieux les évanouissements sélectifs intervenant sur une liaison radiomobile.

1. INTRODUCTION (NU)

Les nouveaux systèmes numériques destinés aux communications radiomobiles civiles ou militaires sont appelés à fonctionner dans des environnements de plus en plus hostiles comme peut l'être, par exemple, un site urbanisé.

Les coupures intervenant lors d'une transmission hertzienne sont dues essentiellement aux phénomènes de trajets multiples présents dans un site urbain où l'existence du signal direct entre l'émetteur et le récepteur n'est pas, en général, assurée. La propagation des ondes est caractérisée par un phénomène de diffraction complexe sur des obstacles mobiles ou non ainsi que par des variations rapides et aléatoires de l'amplitude et de la phase des signaux reçus. Ces phénomènes de propagation se traduisent par des évanouissements sélectifs du signal reçu, en fréquence et en espace, et par un taux d'erreur binaire accru par des interférences intersymboles.

Différents travaux publiés dans la littérature ont démontré que, malgré tous ces inconvénients, il est possible de tirer profit de la présence des multitrajets grâce à la discrimination temporelle et fréquentielle due au large spectre de fréquence produit par l'étalement.

L'étalement de spectre par codage direct permet de discriminer les différents trajets de la propagation en fonction du temps. L'étalement par saut de fréquence, quant à lui, constitue une technique pour lutter contre la sélectivité du canal de transmission. Ces deux principes réduisent notamment l'effet des interférences intersymboles intervenant sur des liaisons en site urbain.

Plusieurs études antérieures ont porté sur le concept de diversité par sélection ou par combinaison qu'offre l'étalement de spectre. Ainsi, PRICE et al [1] et TURIN [2] ont exposé le principe de diversité de multitrajets en considérant l'adaptation du récepteur à la réponse impulsionnelle du canal de transmission à l'aide d'un filtre transversal avec des sorties pondérées et combinées. Par analogie, COOPER et al [3] ont étudié la diversité de fréquence avec une démodulation différentielle sur chacune des branches afin d'éviter les problèmes d'autoadaptativité de la phase.

Cet article présente l'étude et la réalisation d'un système à diversité de multitrajets en prédétection dont le principe est de dissocier chacun des signaux reçus grâce aux propriétés de l'étalement de spectre. L'originalité du système proposé réside dans la réalisation d'un filtre transversal conçu à partir d'une ligne à retard multiprises à ondes acoustiques de surface (SAW) et adapté à la réponse impulsionnelle du canal, obtenue par un convoluteur (SAW).

2. PRESENTATION DE LA LIAISON SANS DIVERSITE (NU)

La liaison numérique à étalement de spectre existante est unilatérale entre une base d'émission fixe et un récepteur mobile; sa structure est montrée à la figure 1.

A l'émission, l'étalement de spectre est réalisé par un codage direct qui consiste à additionner modulo 2 l'information binaire codée par transition, à une séquence pseudo-aléatoire ayant un débit numérique beaucoup plus important que celui du message à transmettre. Puis, le signal subit une modulation de phase à deux états (MDP 2) à la fréquence intermédiaire (FI) avant d'être transposé à la fréquence d'émission (RF), amplifié et émis à partir d'une antenne omnidirectionnelle.

Au niveau du récepteur mobile, le signal large bande est transposé en FI puis attaque un filtre adapté à ondes acoustiques de surface qui permet d'obtenir des pics d'autocorrélation représentatifs de la réponse impulsionnelle du canal et qui autorise ainsi la discrimination des différentes ondes reçues par le récepteur. Une démodulation permet ensuite de récupérer le signal en bande de base.

L'évaluation des performances du système a été effectuée en présence de différentes perturbations pouvant affecter la liaison réelle (bruit blanc, brouilleurs, trajets multiples...) et montre un bon accord avec la théorie [4].

Les mesures de propagation à large bande ont permis de déterminer la réponse impulsionnelle instantanée du canal de transmission et ont mis en évidence les phénomènes de trajets multiples. Un traitement statistique des données recueillies lors de campagnes de mesures sur le site réel a conduit à la modélisation du canal urbain [5].

Il est à noter que la synchronisation du système s'effectue généralement sur le premier trajet, reçu avec une puissance suffisante. En cas d'évanouissement profond, le récepteur perd la synchronisation pendant un temps correspondant à la recherche et la détection d'un autre trajet. Pour lutter contre les effets néfastes de ces évanouissements, des techniques de diversité de multitrajets par sélection ou par combinaison peuvent être employées.

3. DIVERSITE PAR COMBINAISON DES MULTITRAJETS (NU)

3.1. Description générale du système

A l'émission, la séquence pseudo-aléatoire réalisant l'étalement de spectre a une longueur maximale de 127 bits, un débit de 10 Mbits/s et est décrite par le polynôme générateur ($x^7 + x^3 + x^2 + x + 1$). Pour assurer le synchronisme du débit numérique du code et celui de l'information binaire (39,4 Kbits/s), l'horloge servant à élaborer le message numérique est réalisée à partir de la détection, une fois sur deux, du mot de 7 bits à "1" présent dans chaque séquence d'étalement. La modulation à deux états de phase est effectuée à la fréquence intermédiaire de 70 MHz puis l'émission se fait à la radio fréquence de 910 MHz.

A la réception, le dispositif à diversité de multitrajets en prédétection est conçu autour de deux filtres, figure 2. Le premier, fonctionnant à la fréquence FI = 70 MHz, est un filtre à ondes acoustiques de surface adapté au signal émis et qui réalise une opération de convolution entre le signal reçu et le code d'étalement, de longueur 127 bits.

Le second est un filtre adapté au signal de sortie du convoluteur. C'est un filtre transversal conçu à partir d'une ligne à retard multiprise à ondes acoustiques de surface dont les différentes sorties sont pondérées en temps réel afin de réaliser l'adaptation du récepteur aux paramètres du canal de transmission et d'effectuer ainsi une combinaison à gain maximal des multitrajets. Ce procédé a pour effet d'éliminer au mieux les évanouissements sélectifs. La démodulation différentielle en prédétection sur chacune des sorties du filtre transversal est élaborée à partir d'une ligne à retard à transfert de charges (CCD) et permet ainsi de supprimer la phase aléatoire intervenant lors de la transmission des différentes ondes réfléchies.

L'évaluation des paramètres du canal (amplitudes a_k, retards τ_k) servant à la pondération instantanée des sorties du filtre transversal, est obtenue après une détection d'enveloppe de la sortie du convoluteur et un démultiplexage ayant pour effet de transformer les pics de corrélation en signaux dont l'amplitude est proportionnelle à celle des différents pics.

Les signaux ainsi pondérés et démodulés en sortie de la ligne à retard CCD sont combinés. Ensuite, un circuit de décision reconstitue la polarité des bits de l'information numérique transmise. Tout se passe alors comme si le récepteur ne recevait qu'une seule onde ne subissant pas d'évanouissement.

3.2. Filtre adapté

A la réception, le signal en RF = 910 MHz, codé à large bande (10 MHz), est capté par une antenne omnidirectionnelle placée sur le toit du véhicule mobile. Un amplificateur faible bruit et un filtre passe-bande précèdent la transposition à la fréquence intermédiaire de 70 MHz. Ensuite le filtre adapté est mis en œuvre. Celui-ci est réalisé à partir d'un convoluteur à ondes acoustiques de surface adapté à la séquence émise. Il détermine ainsi la réponse impulsionnelle du canal qui se présente sous la forme d'un train de pics de corrélation, chacun étant représentatif d'un trajet reçu, figure 3. La largeur de bande du système permet une discrimination temporelle des trajets de 100 ns et la fenêtre d'observation obtenue est de 12,7 µs. Plusieurs campagnes de mesures dans un site urbain et notamment dans la ville de Rennes, ont confirmé que le retard maximal subi par les réflexions multiples dépassait rarement cette valeur [4]. Il est à noter que le convoluteur conserve la modulation de phase.

3.3. Filtre transversal

Pour tirer profit des trajets multiples, il faut que le filtre transversal soit adapté à une estimation de l'amplitude de la réponse impulsionnelle du canal. Chaque profil de retard doit être corrélé avec cette réponse impulsionnelle qui est l'image du signal par rapport à un miroir perpendiculaire à l'axe des temps.

D'après [6], un filtre est adapté à un signal s(t) si sa réponse impulsionnelle est de la forme: $h(\tau) = \alpha \cdot s(\Delta - \tau)$, où α et Δ sont des constantes arbitraires.

Dans notre étude expérimentale, le filtre transversal est composé d'une ligne à retard multiprise à ondes acoustiques de surface dont les différentes sorties sont pondérées avec des coefficients réels élaborés à partir de tensions proportionnelles aux amplitudes des signaux reçus et d'un combineur linéaire qui effectue une sommation des différents signaux pondérés ayant, auparavant, subi une démodulation différentielle.

Pour utiliser au mieux la diversité de multitrajets, il faudrait que la ligne à retard ait un grand nombre de sorties espacées d'un pas égal à $1/R_c = 100$ ns, R_c étant la largeur de bande du signal émis. Mais, pour des raisons de coût et de disponibilité, la ligne à retard multiprise, conçue pour le simulateur de canal radiomobile large bande [7], est utilisée dans l'élaboration du système à diversité de multitrajets proposé.

Compte tenu des caractéristiques techniques de la ligne, onze retards τ_k de la propagation ont été choisis: 0 - 0,3 - 0,6 - 0,9 - 1,2 - 1,7 - 1,9 - 2,1 - 2,3 - 2,5 - 2,7 µs. Chaque profil de retard obtenu avec le filtre adapté à ondes acoustiques de surface, dure 12,7 µs; ce qui correspond à la durée de la séquence pseudo-aléatoire. Mais, d'après les valeurs des retards choisis, seule une fenêtre de 2,7 µs sera exploitée. En se basant sur les résultats de la caractérisation des canaux radiomobiles [8], cette valeur est jugée suffisante. Le principe de fonctionnement du filtre transversal est montré à la figure 4. La pondération est obtenue en modulant le signal de rang k en sortie de la ligne, par son amplitude estimée $\hat{a}_k$ et s'effectue en même temps sur toutes les sorties de la ligne à retard, pendant une durée de 100 ns environ.

3.4. Signaux de pondération

L'élaboration des coefficients instantanés de pondération est obtenue par une détection d'enveloppe appliquée à la sortie du filtre adapté. Les différents pics de corrélation sont alors démultiplexés. Le nombre et l'espacement temporel des sorties du démultiplexeur sont imposés par les retards de la ligne à ondes acoustiques de surface. Un échantillonneur bloqueur situé sur chacune des 11 sorties du démultiplexeur a pour effet de transformer le pic de corrélation en une tension continue proportionnelle à l'amplitude de ce dernier. Les signaux de commande sont réalisés avec de la logique combinatoire. Le schéma fonctionnel de l'élaboration des coefficients de

pondération est donné à la figure 5.

La concordance dans le temps des profils en sortie de la ligne à retard et de tous les signaux de pondération est impérative pour assurer une parfaite adaptation du filtre transversal, d'où la nécessité d'une référence stable dans le temps correspondant, en général, au trajet direct.

Dans l'étude théorique menée par TURIN [2], chaque profil de retard est référencé au temps $t_0 = 0$, donc au trajet direct reçu. Mais, lors d'une liaison réelle en site urbain, l'existence du signal direct entre l'émetteur et le récepteur mobile n'est pas, en général, assuré. Par contre, dans un environnement suburbain, le problème est moindre car l'existence de l'onde directe au niveau du récepteur est quasi permanente.

En pratique, dans le cas où l'onde directe n'est pas présente, il faudrait estimer à chaque instant la distance entre l'émetteur et le récepteur mobile afin d'estimer le temps absolu de propagation de l'onde. Pour pallier à ce problème, la référence de temps sera obtenue par la détection du premier signal reçu qui détermine ainsi le début de chaque profil de retard; c'est à dire que le fonctionnement du dispositif s'effectue en temps relatif et non pas en temps réel. Donc, les trajets pris en compte pour la diversité s'étalent sur une durée de 2,7 µs à partir du premier trajet détecté.

3.5. Démodulation différentielle

Les différents pics de corrélation, pondérés en sortie de la ligne à retard à ondes de surface, vont subir une démodulation différentielle afin de s'affranchir des problèmes de phase aléatoire intervenant lors de la transmission du message à travers le canal radiomobile. Le principe est de moduler en amplitude chaque pic de corrélation, contenant la porteuse en FI modulée en phase, avec lui-même mais retardé d'un temps correspondant à une période T_m de l'information binaire.

La démodulation est réalisée avec un dispositif à transfert de charges CCD dont le retard τ est imposé par le nombre des registres à décalage et la fréquence de transfert des charges électriques. Sa valeur est donnée par la relation suivante:

$$\tau = \frac{\text{nombre d'étages du registre}}{\text{fréquence de transfert}}$$

Le débit de l'information binaire a été fixé à: $R_m = 39{,}4$ KHz, d'où $T_m = 25{,}4\ \mu s$ $(= 2 . L. T_c = 2 . 12{,}7\ \mu s)$, L étant la longueur du code d'étalement et T_c son débit.

Cette valeur a été choisie d'une part pour assurer le synchronisme du débit numérique du code et celui de l'information binaire et d'autre part d'après les caractéristiques techniques de la ligne à retard à transfert de charges.

Le principe de la démodulation différentielle est montré à la figure 6. Cette démodulation s'effectue à la fréquence de 70 MHz mais, pour respecter les conditions de fonctionnement de la ligne à transfert de charges, l'opération de retard est réalisée à 10 MHz. Un filtrage passe-bas permet d'éliminer la composante à la fréquence double générée par le mélange. L'étude expérimentale a montré que la polarité des pics de corrélation à la sortie du démodulateur, est bien négative pour des bits transmis de niveau logique "1" et positive pour des "0", figure 7.

Il est à noter que les différentes fréquences intervenant dans le système sont générées à partir d'une même source stable afin d'assurer une bonne cohérence entre elles.

3.6. Régénération de l'information

L'amplitude des pics de corrélation démodulés et combinés est comparée à deux seuils. La sortie d'un tel dispositif est à "0" si l'amplitude est au dessus d'une tension positive et égale à "1" si elle est en dessous d'une tension négative. Ensuite, pour limiter les erreurs sur la décision prise, une opération d'échantillonnage doit être effectué selon le rythme $1/T_m$ régénéré localement.

4. <u>CONCLUSION</u> (NU)

Cette communication a présenté l'étude et la réalisation d'un système à diversité de multitrajets en prédétection dont le principe est de dissocier chacun des signaux reçus grâce aux propriétés de l'étalement de spectre et d'effectuer une démodulation différentielle sur chacun d'eux afin de s'affranchir des variations aléatoires de la phase intervenant lors de la transmission.

L'originalité du système proposé réside dans la réalisation d'un filtre transversal conçu à partir d'une ligne à retard multiprises à ondes acoustiques de surface (SAW) et adapté à la réponse impulsionnelle du canal, obtenue par un convoluteur (SAW). Les coefficients de pondération du filtre transversal varient en temps réel afin d'obtenir l'adaptation du récepteur aux paramètres du canal de transmission et de combiner les multitrajets pour éliminer au mieux les évanouissements sélectifs.

5. <u>REFERENCES</u> (NU)

[1]- R. PRICE - P.E. GREEN
"A Communication Technique for Multipath Channels."
Proceedings of the IRE
March 1958. pp. 555-570.

[2]- G.L. TURIN
"Introduction to Spread-Spectrum Antimultipath Techniques and Their Application to Urban Digital Radio."
Proceedings of the IEEE
Vol. 68. n° 3. March 1980. pp. 328-353.

[3]- C.R. COOPER - R. W. NETTLETON
"A Spread-Spectrum Technique for High-Capacity Mobile Communications."
IEEE Trans. on Veh. Tech.
Vol. VT 27. n° 4. November 1978. pp. 264-275.

[4]- G. EL ZEIN - M. SALEHUDIN - A. DANIEL - J.J. BAI - J. CITERNE
"Etude expérimentale d'une liaison numérique radiomobile à étalement de spectre en site urbain."
AGARD: Propagation effects and circuit performance of modern military radio systems with particular emphasis on those employing bandspreading.
17-21 Octobre 1988. Arcueil (France). pp. 31.1 - 31.10.

[5]- M. SALEHUDIN - G. EL ZEIN - J.J. BAI - A. DANIEL - J. CITERNE
"Estimation de la réponse impulsionnelle du canal radiomobile large bande en site suburbain à 910 MHz."
AGARD: Propagation effects and circuit performance of modern military radio systems with particular emphasis on those employing bandspreading.
17-21 Octobre 1988. Arcueil (France). pp. 11.1 - 11.8.

[6]- G.L. TURIN
"An Introduction to Matched Filters."
IRE Transactions on Information Theory
June 1960. pp. 311-329.

[7]- A. DANIEL - G. EL ZEIN - M. SALEHUDIN - J. CITERNE
"Etude et réalisation d'un simulateur de canal radiomobile à large bande."
L'Onde Electrique. Vol. 68. n° 2. Mars 1988. pp. 82-89.

[8]- P.W. HUISH, E. GURDENLI
"Propagation Measurement and Planning Requirements for Digital Cellular Systems."
Proceedings of the Second Nordic Seminar on Digital Land Mobile Radio Communications.
14-16 October 1986. pp 47.1-47.6.

6. REMERCIEMENTS (NU)

Cette étude a été réalisée en l'exécution d'un contrat émanant de la Direction des Recherches Etudes et Techniques (DRET).

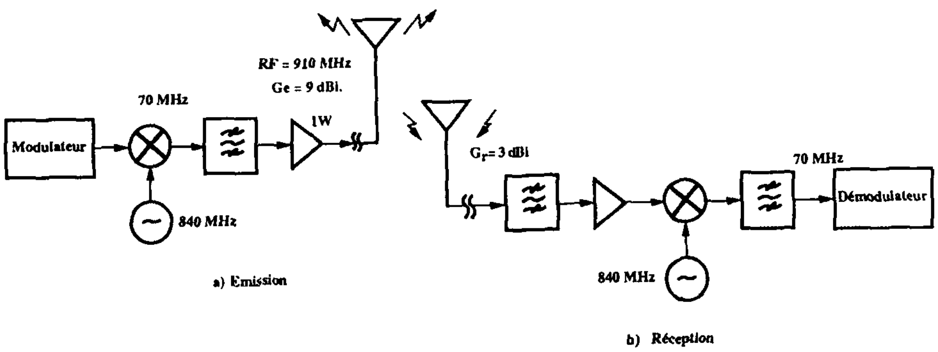


Figure 1: **Liaison numérique à étalement de spectre**

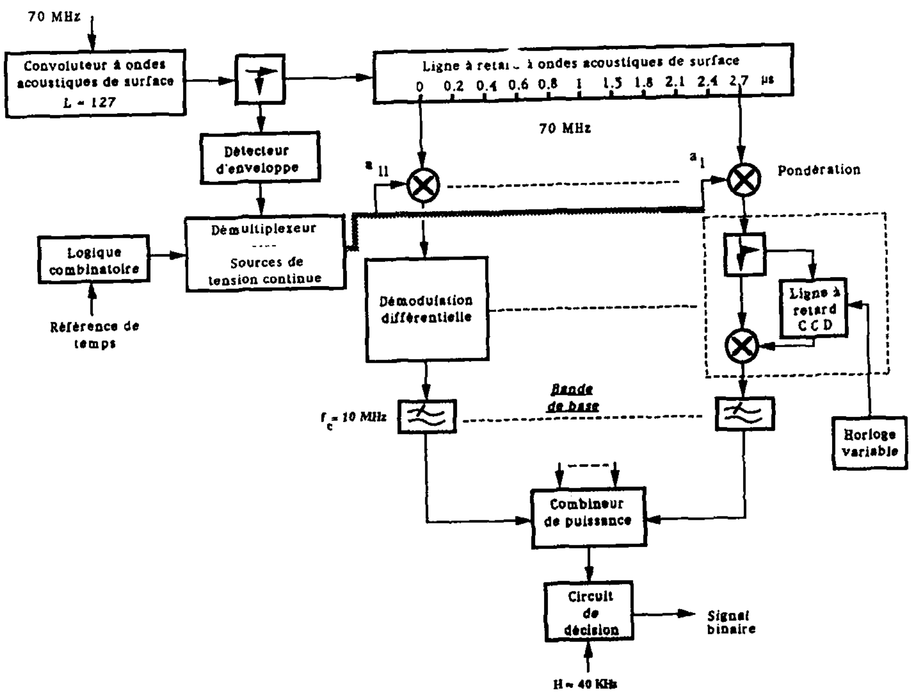


Figure 2: **Diversité par combinaison de multitrajets**
Structure du démodulateur

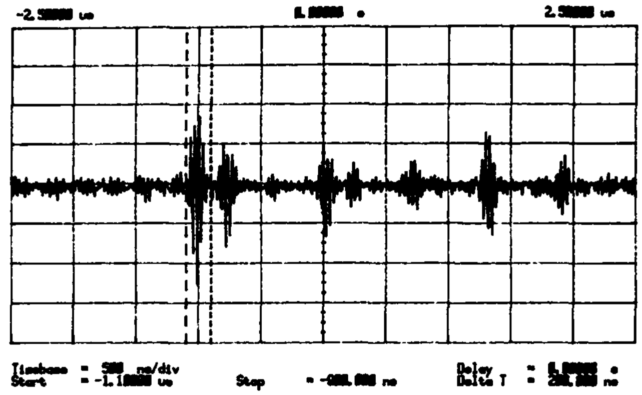
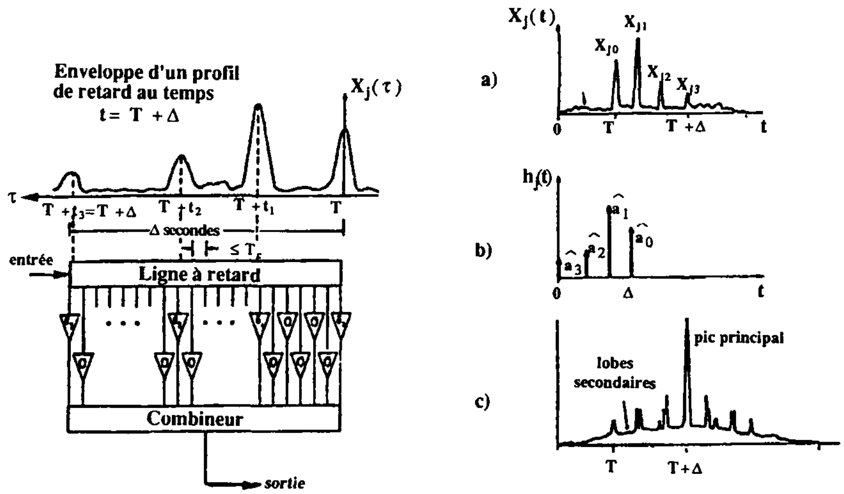


Figure 3: **Profil de retard obtenu à la sortie du filtre adapté**



Filtre transversal

a) Signal d'entrée

b) Réponse impulsionnelle du filtre

c) Signal de sortie obtenu par
la convolutio.ı de a) et b)

Figure 4: **Fonctionnement du filtre transversal**

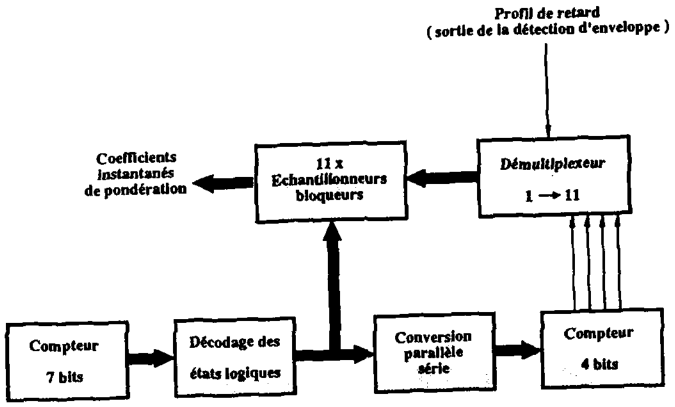


Figure 5: **Elaboration des signaux de pondération**

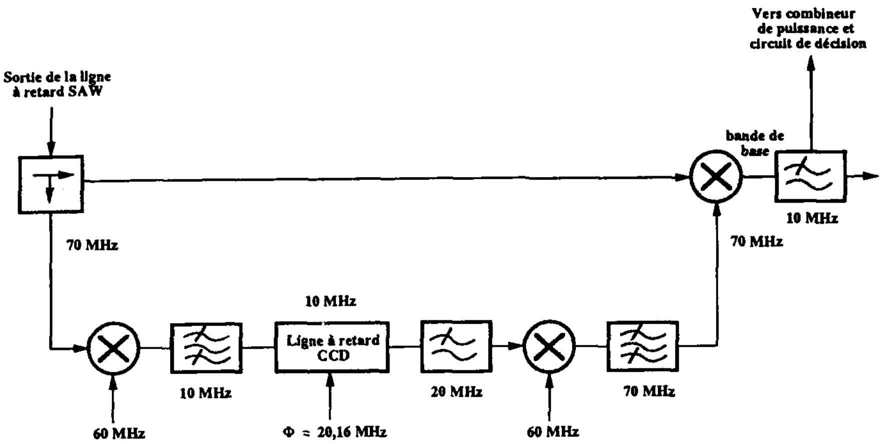
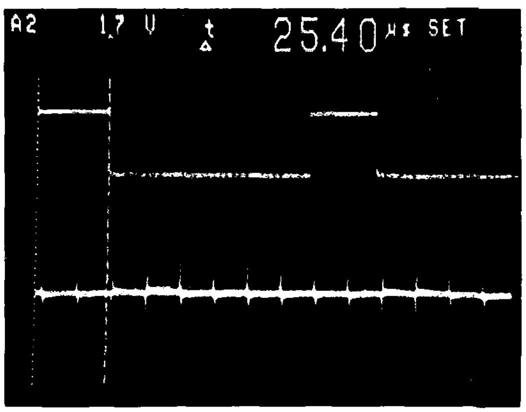


Figure 6: **Démodulation différentielle**



Information binaire

Pics de corrélation en sortie
de la démodulation différentielle

Figure 7: **Signal en sortie de la démodulation différentielle**

DISCUSSION

C. GOUTELARD, FR
English Translation

There is no doubt that your studies are of great interest. In your presentation you use a transverse filter which uses non-uniform distribution multiconnectors, with temporal gaps up to 0.5 microseconds, whereas you specified that they should be 0.1 microseconds. Furthermore, the smallest gaps are at the beginning of the line. It seems to me that this makes your system clearly sub-optimal. Can you explain your choice?

AUTHOR'S REPLY
English Translation

Part of the design process for a wideband mobile radio channel simulator included a surface acoustic wave multiconnector delay line, specially designed to reproduce the different propagation delays encountered in an urban environment. The amount and value of the outputs being limited by the component technology, delays were selected between 0 and 6.5 microseconds with unequal steps, in order to comply with Poisson's law of distribution. The 15 values were as follows: 0 - 0.2 - 0.4 - 0.6 - 0.8 - 1 - 1.5 - 1.8 - 2.1 - 2.7 - 3.5 - 4.5 - 5.5 - 6.5 microseconds.

For reasons of cost and availability, this delay line was used for the design of the transverse filter employed in the multipath diversity system. As was stated in the article, in order to derive maximum benefit from multipaths, the filter must be designed for estimation of the channel's pulse response amplitude. Each delay profile must be correlated with this response, which is the signal image in a mirror perpendicular to the time axis.

According to this theory, delay line outputs are considered from right to left. The last 4 delays being spaced 1 microsecond apart, and diversity achieved on the 11 multipaths whose delays are between 0 and 2.7 microseconds. The values considered are therefore as follows: 0 - 0.3 -0.6 - 0.9 - 1.2 - 1.7 - 1.9 - 2.1 - 2.5 - 2.7 microseconds.

Clearly the system so designed is sub-optimal, as it does not take into account all the propagation paths, but it is nevertheless more efficient than a receiver without diversity.

NEW PERSPECTIVES FOR THE INVERSION OF BACKSCATTER IONOGRAMS

N. RUELLE*, J.Y. LE SAOUT*, L. BERTEL **

* Centre National d'Etudes des Télécommunications
LAB/MER/GER, route de Trégastel
BP 40, 22301 Lannion, France.

** Université de Rennes I
Laboratoire Radiocommunications
Campus de Beaulieu
35042 Rennes Cedex, France.

SUMMARY

HF backscatter sounders offer the perspective of real-time sensing of ionospheric radio propagation conditions in an extended area. However, to this date, difficulties are experienced with the interpretation of sounding data, which places a severe limit on the operational use of these sounders. The French National Telecommunications Research Centre is constructing a sounder of original design. Wc evaluate the new possibilities for data interpretation offered by this system and discuss possible operational implications.

1. INTRODUCTION

HF sky-wave backscatter sounding is a technique which has been used in the past for purposes such as sea-state sensing, ionospheric research and HF over-the-horizon radar (see Anderson & Lees (1) for a review).

Backscatter sounders are in effect radars which use the large scale stratifications of the ionospheric plasma as refracting elements and an extended area of the ground as the target (scatterer).

These systems can be used across the HF band to make a number of measurements on the backscattered echoes (propagation time, signal level, Doppler spectrum). Backscatter sounders offer the possibility of real-time sensing of ionospheric radio propagation conditions in an extended area (a circle of several thousand kilometers radius) around the sounder site. Thus potentially they represent powerful decision aids for the operational exploitation of systems using HF propagation over long distances. This potential could be realised by using soundings to map ionospheric parameters (and their horizontal variations) in real time, enabling the user to predict propagation conditions for any link within the area, using ray-tracing algorithms.

However operational systems require regular and accurate data ; existing methods for interpreting backscatter sounding data give results which are irregular in accuracy (over time) and usually coarse for parameters such as gradients. This severely restricts operational use.

2. THE INVERSION PROBLEM

The interpretation of backscatter sounding data generally makes use of step-frequency backscatter ionograms. These ionograms are defined as the representation of backscattered echo signal level on a frequency/time delay graph.

To derive an electron density profile over the propagation path from an ionogram is known as backscatter ionogram inversion.

The part of the ionogram that is often (but not always) used for this purpose is the so-called leading edge (minimum time delay) curve where a focusing of energy occurs.

The direct problem involves calculating a synthetised leading edge curve using a ray-tracing algorithm and a prescribed ionospheric model. The inverse problem implies determining an electron density profile from an experimental leading edge. An ionospheric model must first be adopted. A realistic large scale model of the ionosphere would require the heights, thicknesses and critical frequencies of the conventional ionospheric layers as well as the horizontal variations (gradients) of these parameters. Possible models are described in Millman et al. (2) and Dyson & Bennett (3).

Because the data (leading edge curve) is not abundant, if the inverse problem is to be reasonably well posed, i.e. giving an unambiguous reconstruction of the ionospheric profile, only a few (most important) parameters can be determined. The remainder must be given a priori values (thus simplifying the model). Generally the aim is to determine the critical frequency and height of the F2 layer at least (because refraction of HF waves mostly takes place in this layer). These parameters are determined at points over the propagation path from successive frequency intervals of the leading edge, because indications of horizontal gradients are particularly valuable.

Since the equations governing the profile to leading-edge transformation are non-linear, the best inversion approach is iterative, with a scheme to adjust the profile parameters so as to minimise the misfit between experimental and synthetised leading edge (for a given frequency interval).

Rao (4) proposes a 3 parameter model, Du Broff et al. (5) a 6 parameter model.

As is common with multiparameter nonlinear inverse problems (Kennett et al. (6)) leading edge inversion as described above is an ill-conditioned problem (in the sense that it is difficult to separate the effect of the various parameters upon the data) ; ill-conditioned problems react very badly to noisy data. Secondary ionospheric parameters (which must be given a priori values in the model) often cannot be predicted accurately from a distance (Hunsucker (7)) because they exhibit considerable spatial variation. Some of these parameters (interlayer "valley" for example) have an effect upon the leading edge, which introduces a form of noise in the problem.

This makes the reliability of the inversion process highly irregular, depending upon the ionospheric state, as is shown by Bertel et al. (8). Hatfield (9) reports reliable success in mapping f_0F2 from backscatter data by reducing the number of model parameters to one, thus simplifying the problem sufficiently for it to be well conditioned.

While f_0F2 is the single most important ionospheric parameter, this would be insufficient for accurate propagation prediction.

Thus it appears that for conventional inversion methods, reliability of inversion cannot be attained unless the model is simplified to the point where it is too coarse for operational use. Since operational systems demand reliable data this severely restricts the use of backscatter sounders as aids for operational systems at this time.

3. IMPROVING INVERSION METHODS

To improve this situation, two methods are presented (which may be used in combination). This aim is to input previously unused information as a regularisation term in the inverse transformation to constrain it. This leads to a better conditioned inverse problem (increased resistance to noisy data, for an increased number of model parameters) thus to an inversion process which is more reliable from the operational point of view.

3.1 Doppler

Backscattered echoes involving refraction off ionospheric layers often are affected by a Doppler shift (of the order of 0.1 Hz) due to ionospheric motions. As shown in (1) this can be used to discriminate between echoes involving different layers by plotting them on a time delay-Doppler map at a given frequency. If time delay-Doppler maps taken at well chosen frequencies are used in addition to the usual step-frequency backscatter ionogram, it will be possible in some cases to avoid multimode contamination in leading edge analysis by separating leading edges traces corresponding to each layer (they are indistinguishable on the ionogram). Then each trace may be treated independently and assigned its own parameters, allowing the use of a more complex ionospheric model.

This will be possible only when interlayer Doppler shifts are nonzero, which is not always the case but occurs sufficiently often to extend the reliability of backscatter inversion. However Doppler analysis implies long integration times and large computing demands (due to FFT calculations) which means that it must be used sparingly in operational systems.

3.2 Use of a narrow beam sounder pattern

Knowledge of elevation angles corresponding to points on the leading edge has been shown to help inversion (Blair et al. (10), Hunsucker (11)). This (narrow-beam) technique requires a sounder with special beam-forming characteristics, but has several advantages. In conventional methods, the frequency of each data point on the leading edge is fixed and group delay time is read, thus giving one experimental value per data point. Then for leading edge synthesis, a differentiation is necessary.

When elevation angle is known for a leading edge point both frequency and group delay time can be read, thus doubling the amount of experimental values per data point. (Also making synthesis calculations easier). This enables separation of the effects of critical frequency and layer height upon the data, resulting in a well conditioned problem, which permits a robust inversion (even with noisy data or inaccurate secondary parameters).

The advantage gained is comparable to that of bistatic soundings where accurate inversion is possible because we have information on fixed ground distance.

We will now discuss an attempt to make use of these two methods.

4. THE LOSQUET BACKSCATTER SOUNDER

CNET (The French National Telecommunications Research Centre) is constructing a backscatter sounder at Lannion (Brittany). This is an omnidirectional system, with co-located transmit and receive arrays ; the cyclic rate is 1/2. PSK modulated sequences are transmitted using a pseudo-random coding. At reception these sequences are retrieved by a correlator to obtain a good signal to noise ratio. Goutelard et al. (12) describe this type of system. The operating frequency may be stepped from 6 MHz to 30 MHz.

The transmitting array is circular with 32 wideband biconic antennas.

The receiving array is circular (concentric) with 64 active antennas.

Array directivity patterns are controlled through digital phase shifters on all antennas (Le Saout & Bertel (13)). Beamforming and steering over 360° in azimuth and 5-50° in elevation is possible.

Several operating modes exist, for example step-frequency, time delay analysis (backscatter ionogram), and fixed frequency, fixed time delay, Doppler analysis.
We show the effect of our sounder array pattern on simulated backscatter ionograms.

5. BACKSCATTER SOUNDING SIMULATIONS

Our aim is to show the effect of the sounder array transmit/receive pattern on the synthetised ionograms for a simple quasi-parabolic ionosphere model concentric with the earth (as defined by Croft & Hoogasian (14)), when the line-of-sight elevation angle of the main lobe is varied.

Ray-tracing for the simulations is performed for a quasi-parabolic F2 layer with the following characteristics : critical frequency 6 MHz, base height 200 km, half- width 50 km.

For signal level calculation the approach of (3) is followed : within the framework of geometrical optics we take into account focusing/defocusing of the rays, ionospheric attenuation and sounder pattern (given by the beam-shaping algorithm).

For each elevation line-of-sight angle we display the sounder transmit-receive pattern (showing beamwidth and sidelobe details) and a step-frequency ionogram.

Elevation L.o.s. angle	Sounder pattern	Ionogram
15°	Fig. 2	Fig. 3
25°	Fig. 4	Fig. 5
35°	Fig. 6	Fig. 7

these figures should be compared to figure 1 which shows time delay/frequency calculated for successive elevation angles with the same ionospheric model. This effect of the sounder pattern may then be described.

Cutoffs can be seen on the backscatter ionograms which correspond in elevation angle values to the nulls on the sounder pattern, as we show :

The first null angle sets the level of the lowest (or main lobe) cutoff. If we read the maximum frequency F_{max} and the time delay T_{max} corresponding to the main lobe cutoff at the leading edge of the ionogram then the first null angle $e_{min} = e(F_{max}, T_{max})$ may be read off figure 1.

We have :	e(deg)	F_{max}(MHz)	T_{max}(km)	e_{min}(deg)
	15	22	4000	1
	25	18	2600	5
	35	12	1500	15

On the sounder pattern these values of e_{min} correspond to the first null (at frequency F_{max}). Likewise the effect of secondary sidelobes and nulls is visible (fig. 7). These findings suggest an inversion method

6. A SUGGESTED INVERSION METHOD

A method which we will try to validate using the CNET sounder is the following :

By using Doppler analysis it is hoped that for a given configuration of the sounder pattern, the traces corresponding to each mode can generally be separated. Then each stratification of the ionosphere which contributes a trace may be analysed separately from the lowest to the highest (contribution from the undermost layers may then be included as a correction in the analysis of the topmost).

For each trace, layer parameters are determined knowing the angle e_{min} and using an inversion method based on this knowledge such as described in (8) or by Ruelle et al. (15), which should give a robust inversion.

This should enable us to generate an ionospheric profile with increased reliability (How substantial this increase will be is what we will try to establish).

Mode of operation : if this process can be repeated for several elevation angles then we should have an idea of possible horizontal gradients in the line-of-sight direction . This can be used in an a posteriori correction loop.

It should then be possible to generate maps of ionospheric parameters by rotating the line-of-sight in azimuth. These maps may then be compared with :

- vertical soundings
- prediction maps

for validation.

7. CONCLUSION - OPERATIONAL USEFULNESS

Using a sounder of novel design we will try to validate an inversion method which, would solve some reliability problems of backscatter ionospheric sounders. It would then be possible to generate maps of several ionospheric parameters in a large area. Thus making it possible to have accurate real-time propagation predictions in this area for links which are not equipped with passive or active sounding equipment.

REFERENCES

1 . Anderson S.J. and Lees M.L.: 'High-resolution synoptic scale measurement of ionospheric motions with the Jindalee sky- wave radar', Radio Science, 23-3, pp 265-272 (1988).

2 . Millman G.H. and Bower C.A.: 'An ionospheric model for HF sky-wave backscatter radar', AGARD Conf. Proc., Munich (1988).

3 . Dyson P.L. and Bennett J.A.: 'A model of the vertical distribution of the electron concentration in the ionosphere and its application to oblique propagation studies', Journal of Atm. and Terr. Phys., 50-3, pp 251 -262 (1988).

4 . Rao N.N.: 'Inversion of sweep-frequency sky-wave backscatter leading edge for quasiparabolic ionospheric layer parameters', Radio Science, 9-10, pp 845-847 (1974).

5 . Du Broff R.E., Rao N.N. and Yeh K.C.: 'Backscatter inversion in spherically asymetric ionosphere', Radio Science, 14-5, pp 837-841 (1979).

6 . Kennett B.L.N., Sambridge M.S., Williamson P.R.: 'Subspace methods for large inverse problems with multiple parameter classes', Geophysical Journal, 94, pp 237-247 (1988).

7 . Hunsucker R.D.: 'Characteristic signatures of the midlatitude ionospheric observed with a narrow-beam HF backscatter sounder, Radio Science, 5-6, pp 535-548 (1971).

8 . Bertel L.E., Cole D.G. and Fleury R.: 'The inversion of backscatter ionograms', IPS technical report TR-88-03 (1988).

9 . Hatfield V.E.: 'Derivation of ionospheric parameters from backscatter data', AGARD Conf. Proc. n° 49, 16 (1969).

10 . Blair J.C., Hunsucker R.D. and Tveten L.H.: 'Ionospheric mapping by backscatter', AGARD Conf. Proc. n° 49, 17 (1969).

11 . Hunsucker R.D.: 'Analysis of a backscatter signature obtained with a high-resolution HF radar system', Radio Science, 6-8, pp 763-768 (1971).

12 . Goutelard C., Caratori J. and Joisel A.: 'Technique de sondage électromagnétique par échantillonnage spatio-fréquentiel des signaux de réception, application à l'étude des inhomogénéités du plasma ionosphérique', AGARD Conf. Proc. n° 244, 13 (1977).

13 . Le Saout J.Y. and Bertel L.: 'Les systèmes d'aériens utilisés par le radar HF à rétrodiffusion du CNET', JINA 88 Conf. Proc., pp 509-513 (1988).

14 . Croft T.A. and Hoogasian H.: 'Exact ray calculations in a quasi-parabolic ionospheric with no magnetic field', Radio Science, 3-1, pp 69-74 (1968).

15 . Ruelle N., Bertel L. and Cole D.G.: 'An empirical approach to the inversion of backscatter ionograms', Radio Science (submitted to).

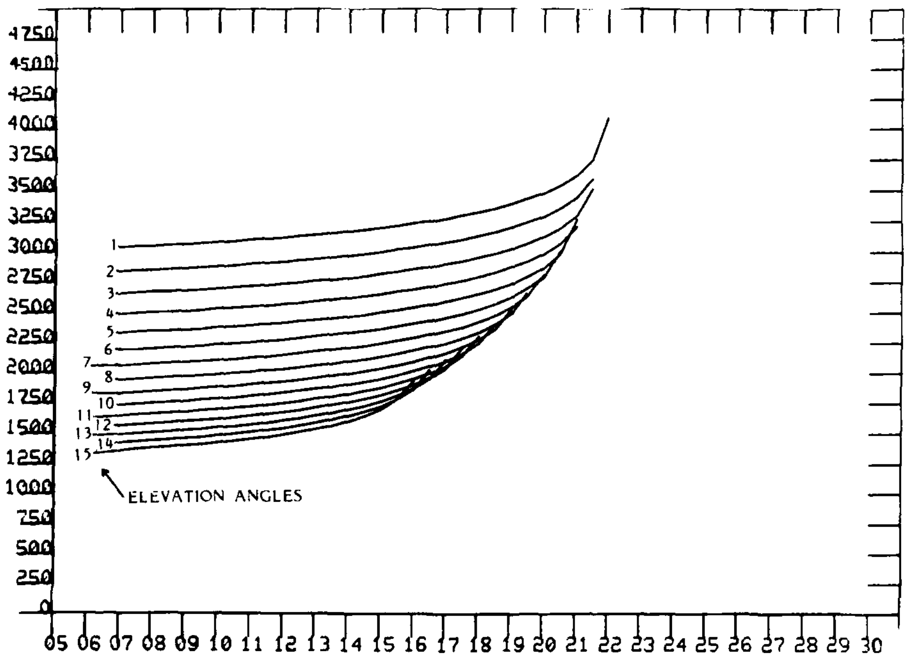


Figure 1 : Time delay (km) as a function of frequency (MHz) for successive ray elevation angles (1° to 15°).

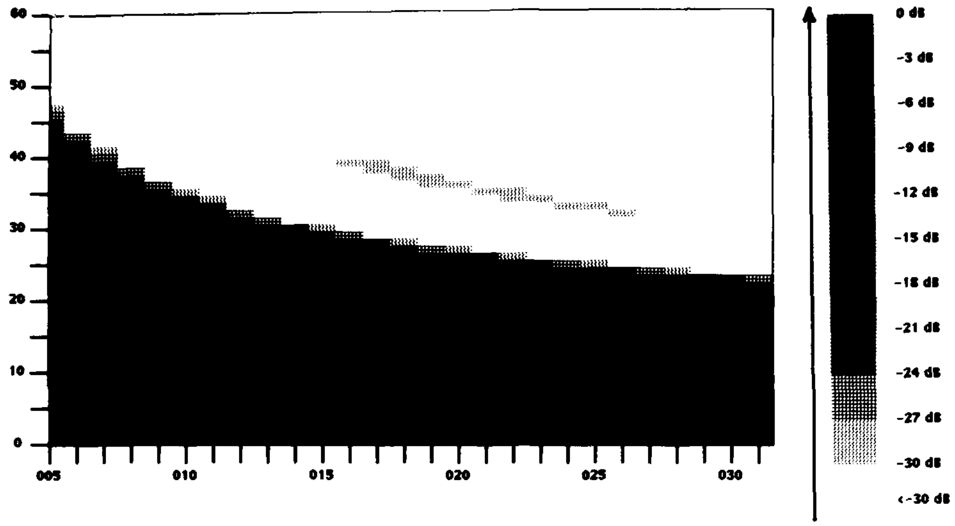


Figure 2 : Sounder pattern ; transmit x receive gain (in dB) as a function of elevation (°) and frequency (MHz) for an elevation line-of-sight angle of 15°.

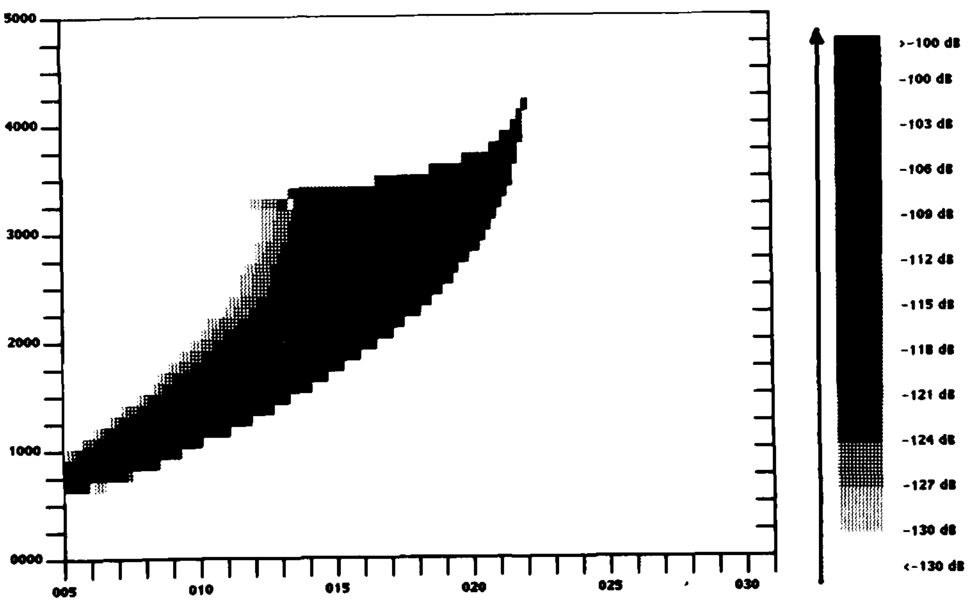


Figure 3 : Synthetised ionogram for the sounder pattern of fig. 2, time delay in km, frequency in MHz.

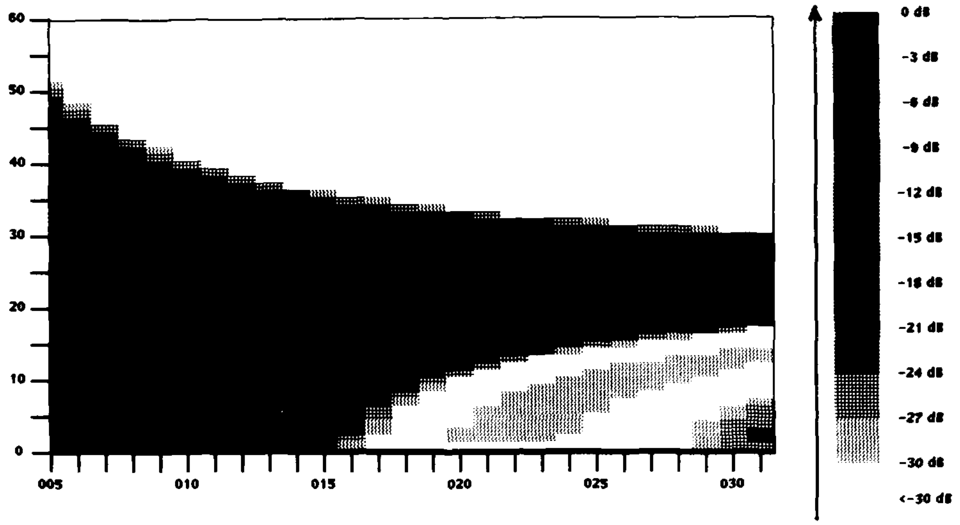


Figure 4 : Sounder pattern for an elevation line-of-sight angle of 25°.

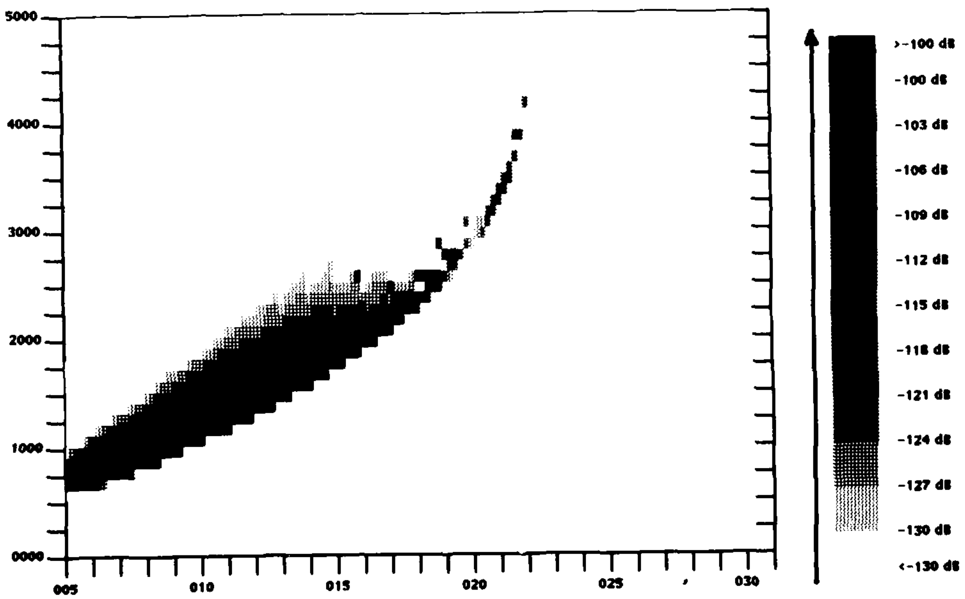


Figure 5 : Synthetised ionogram for the sounder pattern of fig. 4.

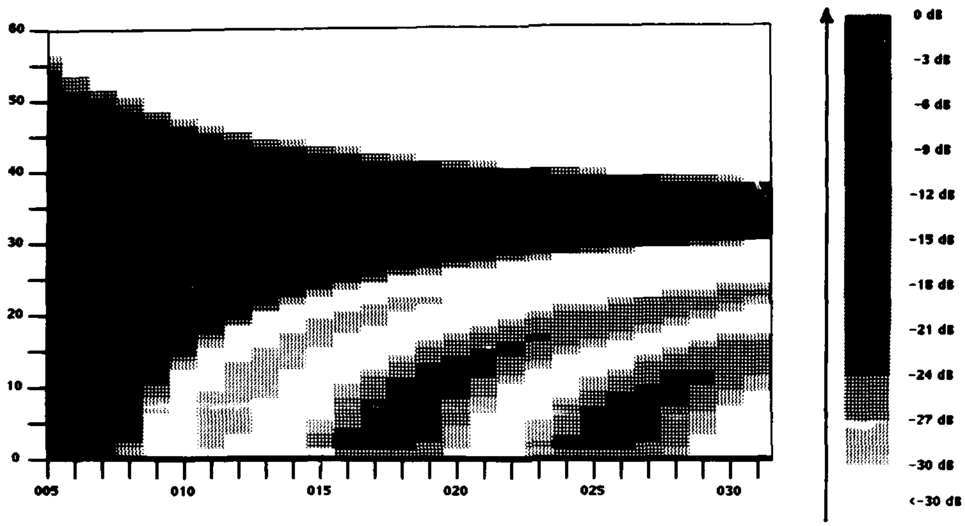


Figure 6 : Sounder pattern for an elevation line-of-sight angle of 35°.

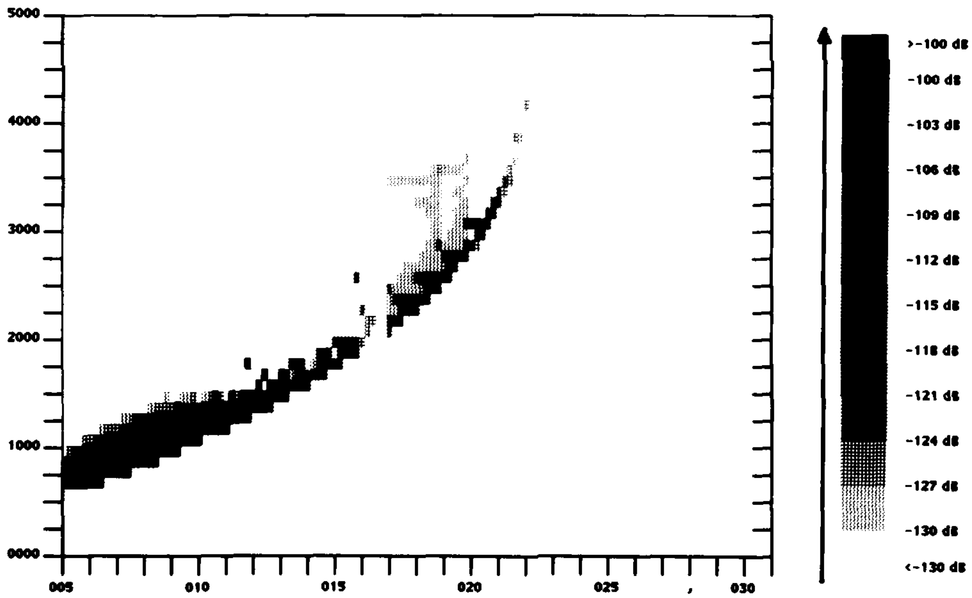


Figure 7 : Synthetised ionogram for the sounder pattern of fig. 6.

DISCUSSION

C. GOUTELARD, FR
English Translation

The idea of using measurement of the incident angles for the inversion of backscatter ionogrammes was raised in the 1960's. To my knowledge, the experimental difficulties were such that nobody experimented with this process. I think your approach is interesting because you equipped yourself with experimental facilities in order to test the method. However, did you evaluate the gain you would achieve as a result of the precision of your measurements? Furthermore, I think you would achieve an additional gain by using azimuthal scanning of the ionosphere, as we do ourselves.

AUTHOR'S REPLY
English Translation

The gain we will achieve will only be precisely evaluated when we have an appropriate data base for the experimental data. The purpose of the presentation was, however, not to show that we increase the precision, but that we reduce the ambiguity of the inversion by using angular data.

K. YEH, US

I have the impression that rays corresponding to the leading edge in a backscattered ionogram are reflected from a narrow height region below the ionospheric peak. If so, there may exist difficulty in deducing ionospheric parameters such as f_oF2 from backscattered ionograms. Please comment on this.

AUTHOR'S REPLY

It does seem that the rays corresponding to the leading edge are reflected just below the layer peak. However, this does not seem to me to preclude the fact that the leading edge depends in a sensitive fashion on the parameters f_oF2 and H_mF2. Certainly in simulations it does. I think that these parameters may then be deduced. It could be necessary, though, to calibrate this determination using vertical (ionosonde) data.

DESIGN OF A COMPENSATION FILTER FOR A QUADRATIC PHASE COMMUNICATION CHANNEL

K. C. Yeh and R. L. Mickelson
University of Illinois at Urbana-Champaign
Urbana, Illinois 61801 USA

SUMMARY

There has always been great interest in understanding and describing the dispersive effects that occur in many types of communication channels. In this paper an analysis of a general dispersive channel is performed by assuming a known transfer function. It is shown that, for bandlimited signals, the channel can be simulated by a set of weighted delay lines. This model clearly illustrates that mitigation of dispersive effects can be achieved via an inverse filtering procedure. Specifically, a compensation filter is designed using a conjugate delay line. Ideally, complete compensation can be achieved and a replica of the transmitted signal can be restored. In practical implementation and application, these idealized conditions can never be met. Nevertheless, this channel simulator and compensation filter are implemented for a quadratic phase transfer function. The system is tested for three basic signals: Gaussian pulse, rectangular pulse, and sudden phase-switched signal. In addition, the case of transmitting multiple pulses is considered. The simulation yields good results for moderate dispersion. As dispersiveness increases, the number of taps in the compensation filter must be increased in order to restore the received signal to a reasonable approximation of its original shape.

INTRODUCTION

An electromagnetic signal propagating through the ionosphere will experience dispersive effects. In addition, other effects arising from the nature of the ionospheric profile and the propagation geometry can also play an important role. These effects have been investigated both theoretically and experimentally for narrow band signals. However, as of late, wideband signal propagation has been of increasing interest to several groups [1]-[3]. Since experimental measurements of system performance over actual ionospheric links are expensive and difficult to obtain, for both narrow and wideband signals, some attention has been placed on the development of channel simulators [4], [5]. A proposal has been made by Nessenbergs [6], to extend the Watterson model [5] from narrow band to wideband signals of 1 or 2 MHz bandwidth.

The purpose of this paper is to further investigate the characteristics of wideband signal propagation through the ionosphere. First, the ionospheric transfer function will be examined and a simulation model will be developed. From this model, a method to mitigate distortion in the received signal will be proposed. More specifically, a compensation filter will be designed and tested assuming a quadratic phase ionospheric transfer function.

1. IONOSPHERIC TRANSFER FUNCTION

To begin the analysis, assume the transmitted is given by

$$s_{Tx}(t) = a_{Tx}(t)e^{j\omega_c t} \tag{1}$$

where a_{Tx} is the complex amplitude of the pulse and is assumed to be slowly varying in time compared to the variation indicated by the carrier angular frequency ω_c. Taking the Fourier transform of both sides of (1) yields

$$S_{Tx}(\omega) = A_{Tx}(\Omega) \tag{2}$$

where S_{Tx} and A_{Tx} represent the spectrums of s_{Tx} and a_{Tx}, respectively, and where $\Omega = \omega - \omega_c$ is the sideband angular frequency centered at the carrier angular frequency.

For a given ionospheric profile and communication circuit, the propagation characteristics of the ionospheric reflection communications channel can be lumped into a transfer function $H(\omega)$. This can be expressed as

$$H(\omega) = |H(\omega)|e^{-j\phi(\omega)} \tag{3}$$

Outside of focussing and caustic regions, the magnitude of the transfer function, $|H(\omega)|$, is slowly varying when compared to the phase, $\phi(\omega)$. In this case, the transfer function of ω is dominated by the phase variation across the band. To examine this phase variation it is convenient to expand the phase term as

$$\phi(\omega) = \phi_c + \tau_c\Omega + \psi(\Omega) \tag{4}$$

where $\phi_c = \phi(\omega_c)$ is the phase at the carrier frequency, $\tau_c = d\phi_c/d\omega_c$ is the group delay for the communication circuit evaluated at the carrier frequency and $\psi(\Omega)$ is the remainder term that depends on Ω nonlinearly. As will be shown later, both the constant phase shift, ϕ_c, and the linear phase shift term, $\tau_c\Omega$, have physical interpretations. For this reason, it is desirable to recast the transfer function as

$$H(\omega) = H_\tau(\Omega)e^{-j(\phi_c+\tau_c\Omega)} \tag{5}$$

where

$$H_\tau(\Omega) = |H(\omega_c+\Omega)|e^{-j\psi(\Omega)} \tag{6}$$

Note that H_τ will be referred to as the time-delayed transfer function.

Next, the impulse response of the system may be defined by utilizing the notation developed above. Let the inverse Fourier transform of H be the impulse response, $h(t)$, and let the Fourier inversion of H_τ be the time-delayed impulse response, $h_\tau(t)$. By applying the inverse Fourier transform to both sides of (5), the following relation can be shown

$$h(t) = h_\tau(t-\tau_c)e^{j(\omega_c t-\phi_c)} \tag{7}$$

With the impulse response specified, the output of the channel may be derived. Letting s_{Rx} represent the received signal, the corresponding amplitude spectrum, $S_{Rx}(\omega)$, is given by

$$S_{Rx}(\omega) = H(\omega)S_{Tx}(\omega) \tag{8}$$

Substituting (2) and (5) into (8) and taking the inverse Fourier transform yields the following expression for the received signal

$$s_{Rx}(t) = a_{Rx}(t)e^{j(\omega_c t-\phi_c)} \tag{9}$$

where the complex amplitude of the received signal is given by

$$\begin{aligned} a_{Rx}(t) &= \frac{1}{2\pi}\int_{-\infty}^{\infty} H_\tau(\Omega)A_{Tx}(\Omega)e^{j\Omega(t-\tau_c)}d\Omega \\ &= a_{Tx}(t-\tau_c) * h_\tau(t) \end{aligned} \tag{10}$$

Thus, the received complex amplitude is simply the time-delayed transmitted complex amplitude convolved with the time-delayed impulse response of the communication channel.

2. CHANNEL SIMULATION AND COMPENSATION

There have been efforts in describing the propagation of pulses in a strongly dispersive medium, e.g., [7]. While it is possible to introduce an ordering small parameter, based on which a set of equations can be obtained, the hierarchy must be solved iteratively to produce, in principle, a result accurate to high orders. Even if the desired received complex amplitude can be computed, the computational procedure is not easily applied in the area of communication system simulation and of finding methods to mitigate pulse distortion. Thus, in this paper, the problem will be approached by attempting to model the dispersive channel effects by a set of delay lines, where each delay line is weighted differently by a tap gain function. This modeling technique has been utilized by several authors [4]–[6].

With this motivation, the received complex amplitude given by (10) can be re-examined. As is generally the case, the transmitted amplitude spectrum $A_{Tx}(\Omega)$ is bandlimited, say from $-\Omega_0/2$ to $\Omega_0/2$. Even if $A_{Tx}(\Omega)$ is not bandlimited, it will be assumed that almost all the signal energy is concentrated within a bandwidth of Ω_0 and that truncation of the spectrum outside of this bandwidth will not introduce detrimental effects on the pulse shape. With this observation. the time-delayed transfer function, H_τ, can be represented by a Fourier series or

$$H_\tau(\Omega) = \sum_{m=-\infty}^{\infty} H_m e^{jmT_0\Omega} \qquad ; |\Omega| < \frac{\Omega_0}{2} \tag{11}$$

where $T_0 = 2\pi/\Omega_0$. The Fourier coefficients, H_m, can be found from

$$H_m = \frac{1}{\Omega_0}\int_{-\frac{\Omega_0}{2}}^{\frac{\Omega_0}{2}} H_r(\Omega)e^{-jmT_0\Omega}d\Omega \tag{12}$$

Using this representation for H_r, the received complex amplitude given by (10) reduces to

$$a_{Rx}(t) = \sum_{m=-\infty}^{\infty} H_m a_{Tx}(t - \tau_c + mT_0) \tag{13}$$

Thus, the received amplitude can be expressed as a superposition of the input amplitudes variously delayed and weighted. This equation leads to the tapped delay line channel model mentioned above. A schematic representation of this channel is shown in Figure 1. In this model the delay is centered about the group delay of the communication circuit, τ_c. The deviation from τ_c is equal to the integer multiples of $T_0 = 2\pi/\Omega_0$ which is usually small in comparison to τ_c. The Fourier coefficients, H_m, represent the tap-gain functions.

It is important to emphasize that the expression given by (13), and subsequently, the channel model, are exact. However, the equation and model become approximate if either the series is truncated at a finite m, H_r is not bandlimited or H_r is only approximately known.

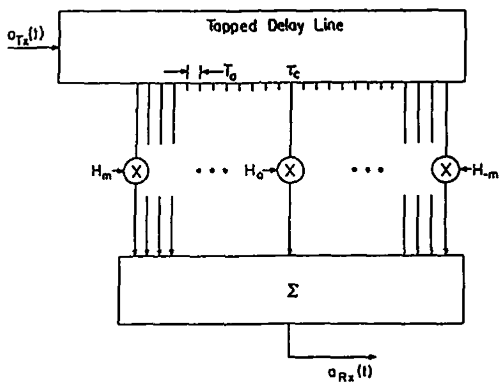


Figure 1. Tapped delay line model of communcations channel.

With the development thus far, it can be seen that to mitigate the pulse distortion which occurs in the channel, a compensation filter can be utilized. One method of mitigation is to place the filter at the receiver, as shown in Figure 2. Then, if the filter has a time-delayed transfer function of $F(\Omega)$ which is equal to $H_r^{-1}(\Omega)$ then the filter output, $\bar{a}_{Rx}(t)$, will have the distortion produced by the dispersive channel completely mitigated.

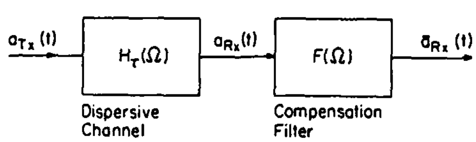


Figure 2. Distortion and compensation represented schematically as filters.

As frequently happens for ionospheric reflection channels, the phase dominates the characteristics of the transfer function. In this case, the time-delayed transfer function, originally given by (6) can be simplified to

$$H_\tau(\Omega) = |H|e^{-j\psi(\Omega)} \tag{14}$$

where $|H|$ is a constant across the band. For this situation, the compensation filter can be simplified to equal $H_\tau^*(\Omega)$. Again, for bandlimited signals $H_\tau^{-1}(\Omega)$ can be represented as a Fourier series over the band $|\Omega| < \Omega_0/2$. This leads to the following expression describing the transfer characteristics of the compensation filter.

$$\bar{a}_{Rz}(t) = \sum_{m=-\infty}^{\infty} H_m^* a_{Rz}(t - mT_0) \tag{15}$$

The resulting compensation filter is shown in Figure 3 where an arbitrary delay, τ_d, has been introduced to make the filter realizable. The compensation filter shown in Figure 3 may be known as a conjugate delay line. Note that, as with the channel model, this compensation network is only ideal if an infinite number of taps are utilized. The compensated output from this ideal filter would then be

$$\bar{a}_{Rz}(t) = |H|^2 a_{Tz}(t - \tau_c - \tau_d) \tag{16}$$

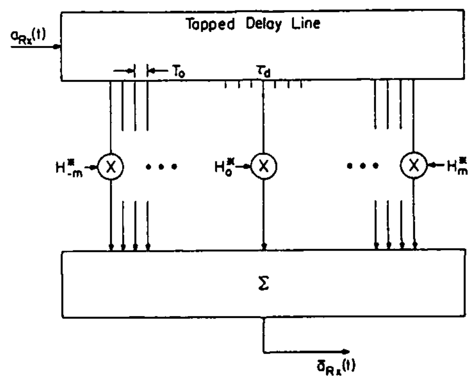


Figure 3. Tapped delay line model of compensation filter.

3. SIMULATION RESULTS

To test the theories just presented a channel transfer function must be specified. Though the theories presented apply to a general transfer function, for computational purposes, the communications channel will be assumed to be a quadratic phase channel for the remaining analysis. For ionospheric applications, the quadratic phase term seems to provide the major dispersive effects and has been utilized in several other investigations [8], [9].

Therefore, the time-delayed transfer function will be assumed to be of the form

$$H_\tau(\Omega) = \exp\left(-j\frac{\tau_c'\Omega^2}{2}\right) \tag{17}$$

where $\tau_c' = d^2\phi_c/d\omega_c^2$. Note that τ_c' is a quantity which can be obtained from the oblique ionogram by measuring the slope of the group delay τ_c as a function of frequency. For simplicity, this parameter will be referred to as the dispersion coefficient.

The next step in development of a compensation filter is to compute the Fourier coefficients of H_τ via (12) to obtain

$$H_m(\Delta) = \frac{1}{2\Delta\sqrt{\pi}} e^{j\frac{m^2}{2\Delta^2}} \int_{\frac{m}{\Delta\sqrt{\pi}}-\Delta\sqrt{\pi}}^{\frac{m}{\Delta\sqrt{\pi}}+\Delta\sqrt{\pi}} e^{-j\frac{\pi u^2}{2}} du \tag{18}$$

where by definition $\Delta = \sqrt{\tau_c'}/T_0$ is a parameter indicating the importance of dispersion over the bandwidth. Note that for this quadratic-phase channel $H_{-m} = H_m$.

Before continuing some notation should be introduced. The compensation filter can only contain a finite number of taps and thus, let the filter output be denoted

$$\bar{a}_{Rx}^{(M)}(t) = \sum_{m=-M}^{M} H_m^*(\Delta) a_{Rx}(t - mT_0) \tag{19}$$

where $2M + 1$ taps exist in the filter model. Also, the dispersiveness of the channel must be described. A parameter, $\delta = \sqrt{\tau_c'}/T_{Tx}$ where T_{Tx} is related to transmitted pulse duration, will be used to describe the importance of dispersion. The signal will be essentially undistorted when $\delta \ll 1$, while the distortion is very important if δ is comparable to or greater than unity.

Next three signals will be utilized to test the compensation filter. The three signals will be the Gaussian pulse, rectangular pulse, and a sudden phase switch signal. In addition a multiple pulse case will be examined.

First consider the transmission of a Gaussian pulse of the form

$$a_{Tx}(t) = \exp\left(\frac{-t^2}{2T_{Tx}^2}\right) \tag{20}$$

Figure 4 shows the results of transmitting this Gaussian pulse through a channel with $\delta = 0.7$ and with $T_{Tx} = 1$. As can be seen in Figure 4(a), the received pulse has been dispersed and, in fact, the received pulse length is about 11 percent longer than the transmitted pulse width. Also, the peak amplitude has been reduced from unity to approximately 0.95. The output of the compensation filter with $M = 50$ is very much similar to the transmitted pulse shape and duration. The peak amplitude has been restored to approximately unity. However, "wings" appear for $|t - \tau_c| > 3T_{Tx}$. By increasing M these wings can be diminished. From Figure 4(b), it can be seen that the phase of the compensated signal very closely follows that of the transmitted signal for times when the amplitude of the transmitted signal is appreciable and only departs from the desired signal for $|t - \tau_c| > 3T_{Tx}$, when the amplitude is very small.

Figure 5 shows the results of transmitting the same Gaussian pulse over a channel with $\delta = 2$. Figure 5(a) shows that the received pulse duration is now more than four times the transmitted pulse width. The peak amplitude has been reduced to one-half the transmitted value. Once again, the mitigated signal has been computed for $M = 50$. The peak amplitude has been increased to almost 0.8, although not quite restored to unity; the pulse has been much narrowed, although not quite restored to the original shape. The phase plot, shown in Figure 5(b) is similar to that with $\delta = 0.7$, except that now the region where the phase has been restored is for $|t - \tau_c| < 2T_{Tx}$, which is narrower than the original case.

Next consider the case of transmitting a rectangular pulse of the form

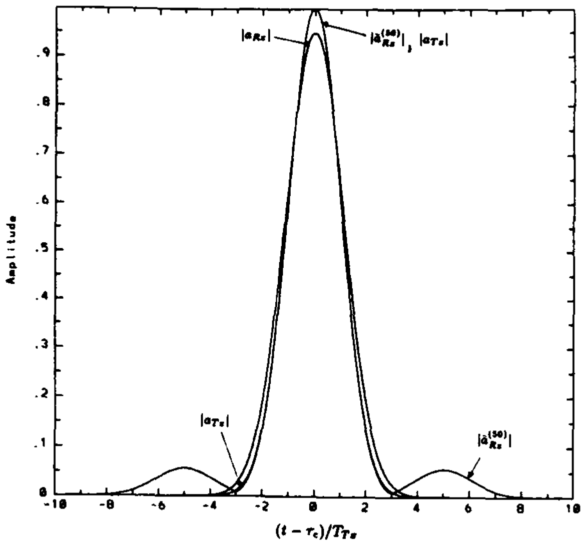
$$a_{Tx}(t) = \text{rect}\left(\frac{t}{2T_{Tx}}\right) = \begin{cases} 1 & ;|t| < T_{Tx} \\ 0 & ;|t| > T_{Tx} \end{cases} \tag{21}$$

where T_{Tx} is set to unity. Transmission of this signal through a quadratic channel with $\delta = 0.7$ yields the results shown in Figure 6. The received signal is severely distorted in amplitude and phase. However, mitigation with $M = 50$ produces a much sharper pulse which begins to exhibit the rectangular pulse shape. Note that ripples are still present though. Figure 7 shows the output of a compensation filter with $M = 200$. The results show that the pulse shape and phase characteristics are greatly enhanced.

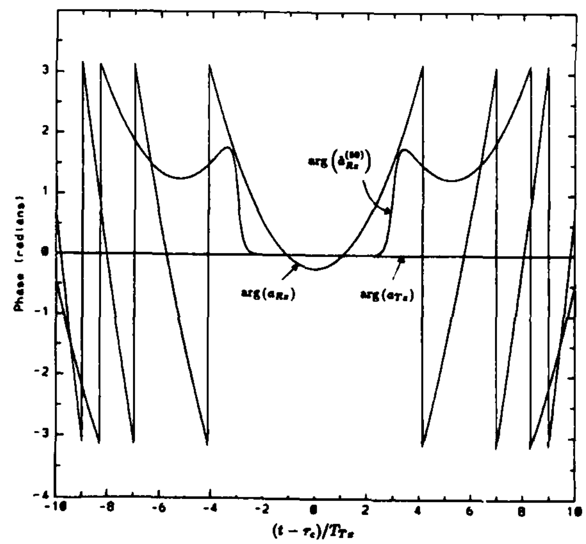
The compensation filter was also tested under the assumption that a sequence of pulses was transmitted. Specifically, the signal

$$a_{Tx}(t) = \text{rect}\left(\frac{t}{2T_{Tx}}\right) - \text{rect}\left(\frac{t - 2T_{Tx}}{2T_{Tx}}\right) \tag{22}$$

was transmitted over the quadratic phase channel with $T_{Tx} = 1$. The transmitted, received and mitigated signals are shown in Figure 8. From (22), the transmitted signal consists of two rectangular pulses; one

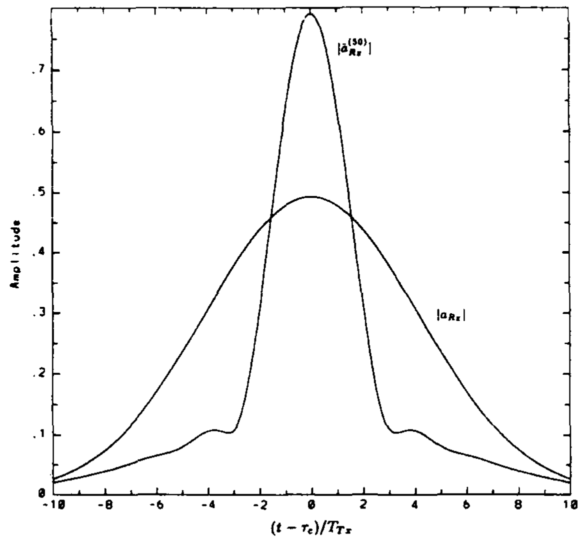


(a)

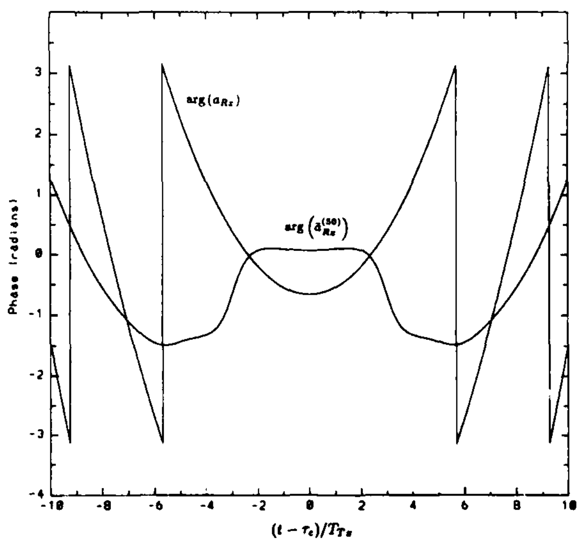


(b)

Figure 4. Transmitted, received, and mitigated Gaussian pulse amplitude and phase ($\delta = 0.7$ and $M = 50$).

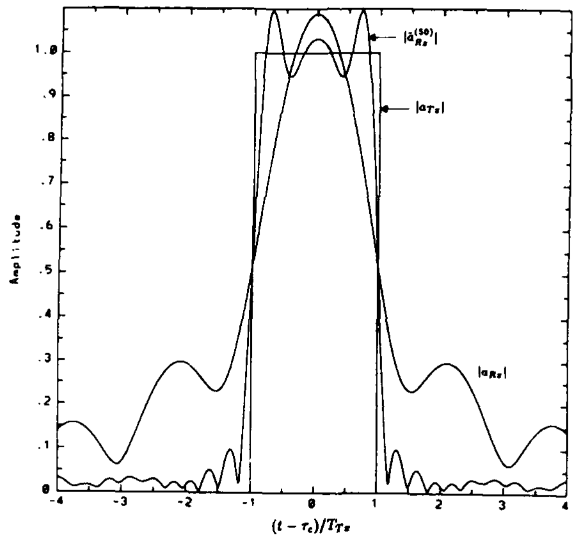


(a)

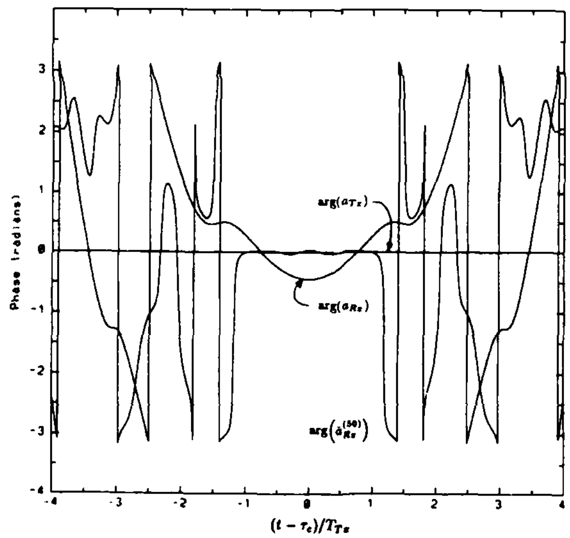


(b)

Figure 5. Received and mitigated Gaussian pulse amplitude and phase (δ = 2 and M = 50).

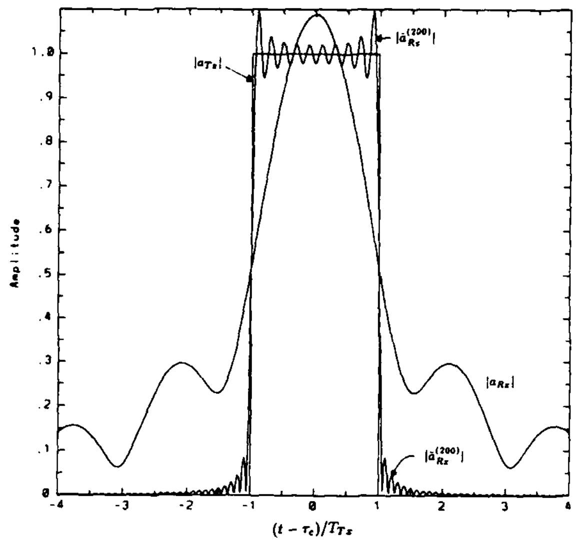


(a)

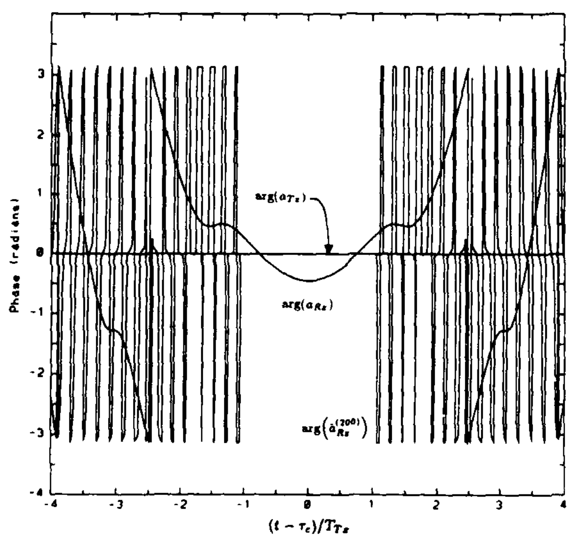


(b)

Figure 6. Transmitted, received, and mitigated rectangular pulse amplitude and phase ($\delta = 0.7$ and $M = 50$).

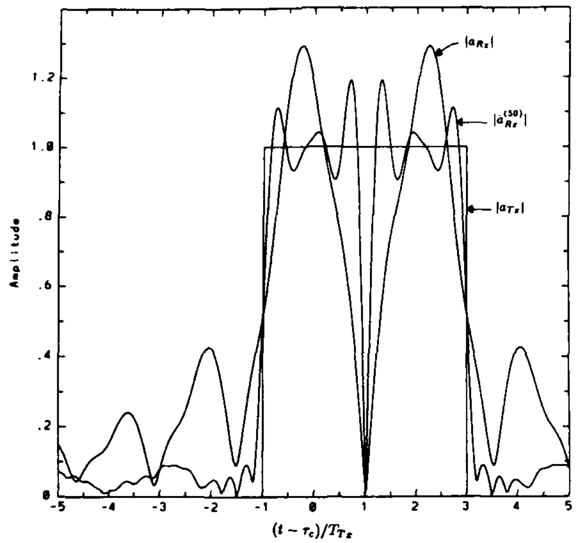


(a)

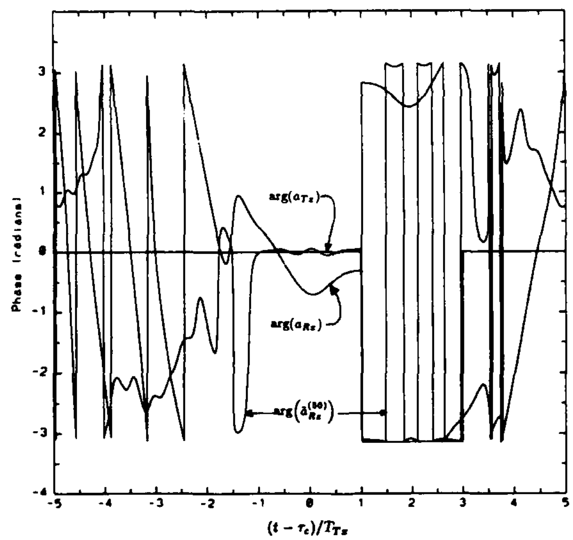


(b)

Figure 7. Transmitted, received and mitigated rectangular pulse amplitude and phase ($\delta = 0.7$ and $M = 200$)).



(a)



(b)

Figure 8. Transmitted, received and mitigated rectangular pulse train amplitude and phase ($\delta = 0.7$ and $M = 50$)).

centered at $t = 0$ and another directly after the first, but with opposite sign. Thus, the desired output should have unity amplitude for two pulse widths and should have zero phase during the first pulse width and a phase of π or $-\pi$ radians during the second pulse width. From the plots, it can be seen that the channel distorts the received train of pulses severely. The dispersive channel tends to expand the pulse duration resulting in one pulse overlapping another. This is often referred to as intersymbol interference. The compensation filter acts to reduce this interference by simply mitigating the channel induced distortion and subsequently, reducing the pulse overlap. As shown in Figure 8, the compensated signal exhibits an improved pulse shape and the phase characteristics are almost ideal (0 for $|t - \tau_c| < T_{Tx}$ and $\pm\pi$ radians for $T_{Tx} < t - \tau_c < 3T_{Tx}$) within the interval, $-T_{Tx} < t - \tau_c < 3T_{Tx}$.

Next, a sudden phase switch signal was transmitted through the quadratic phase channel. This type of signal may be represented by

$$a_{Tx}(t) = \begin{cases} -j & ;t < 0 \\ -je^{j\gamma} & ;t > 0 \end{cases} \tag{23}$$

Examples of these signals are binary phase-shift keying (BPSK) signals where $\gamma = \pi$, and quadriphase-shift keying (QPSK) modulation where $\gamma = \pi/2$. The results of transmitting a BPSK signal over a channel with $\delta = 0.7$ are shown in Figure 9. As with all of the previous cases presented, the received signal distortion is greatly reduced after being filtered by the compensation network.

Lastly, the effect of increasing the number of taps, M, in the compensation filter was examined quantitatively. An error measure was defined as the average mean square error (AMSE) between the transmitted and mitigated pulse amplitude, outside of some predetermined pulse width. To perform the analysis, a rectangular pulse, $\mathrm{rect}(t/2T_{Tx})$, was transmitted, and the error was defined as the average signal power at the filter output in the interval $|t-\tau_c| > T_{Tx}$. The error was calculated for $M = 0$, 50, 100, 150, and 200. Note the error in the case of $M = 0$ is the error present in the unfiltered received signal. Figure 10 shows results for various levels of channel dispersion. As expected, the error tends to decrease rapidly as M increases for low to moderate dispersion, e.g., $\delta = 0.7$ and 1. This is also true for a higher level of dispersion, $\delta = 2$, but is not entirely true for the case $\delta = 3$.

4. CONCLUSIONS

By writing the ionospheric transfer function in terms of a time-delayed transfer function and applying Fourier series techniques, a channel simulator was developed to model pulse dispersion. It was shown that the communications channel could be modeled as a tapped delay line where the channel output was a superposition of replicas of the transmitted complex amplitude variously delayed and weighted.

From the channel model, a mitigation scheme was developed by simply designing a conjugate tapped delay line to compensate for the original pulse distortion. To test this mitigation scheme, a compensation filter was designed assuming a quadratic phase communications channel. Using three different signals, the compensation technique was tested and substantial improvement was obtained after mitigation when the channel was moderately dispersive.

In our analysis, two important assumptions have been made. In the first assumption, the transfer function $H(\omega)$, as given in (3), is understood to be time independent. In the second assumption, $H(\omega)$ has a constant magnitude and a phase up to a quadratic term in sideband in a so-called quadratic phase channel. The time-delayed transfer function is then given by (17). In reality, of course, neither of these two assumptions is valid. Experimentally, the magnitude and phase of the transfer function are found to be both functions of ω and t. Identical mathematical analysis can be carried out for such a time varying channel, except now the transfer function must be approximately stationary in comparison to the total tap length in the compensation filter. For example, for a pulse of length 1 μs and a filter of 101 taps (corresponding to $M = 50$), the ionospheric channel is required to be approximately stationary for a time of 0.1 ms. This is not such a stringent requirement as the ionospheric fading is on the order of seconds. Even under spread-F conditions, the fading may rise up to 10 ms which is still small when compared to 0.1 ms. Under this quasi-stationary approximation, the compensation filter coefficients must be computed from $F(\Omega, t) = H_\tau^{-1}(\Omega, t)$. The resulting coefficients are functions of time and must be updated as the channel characteristics vary. The problems of how frequently must the updating be performed and how important are small deviations in filter coefficients must be left for future investigation.

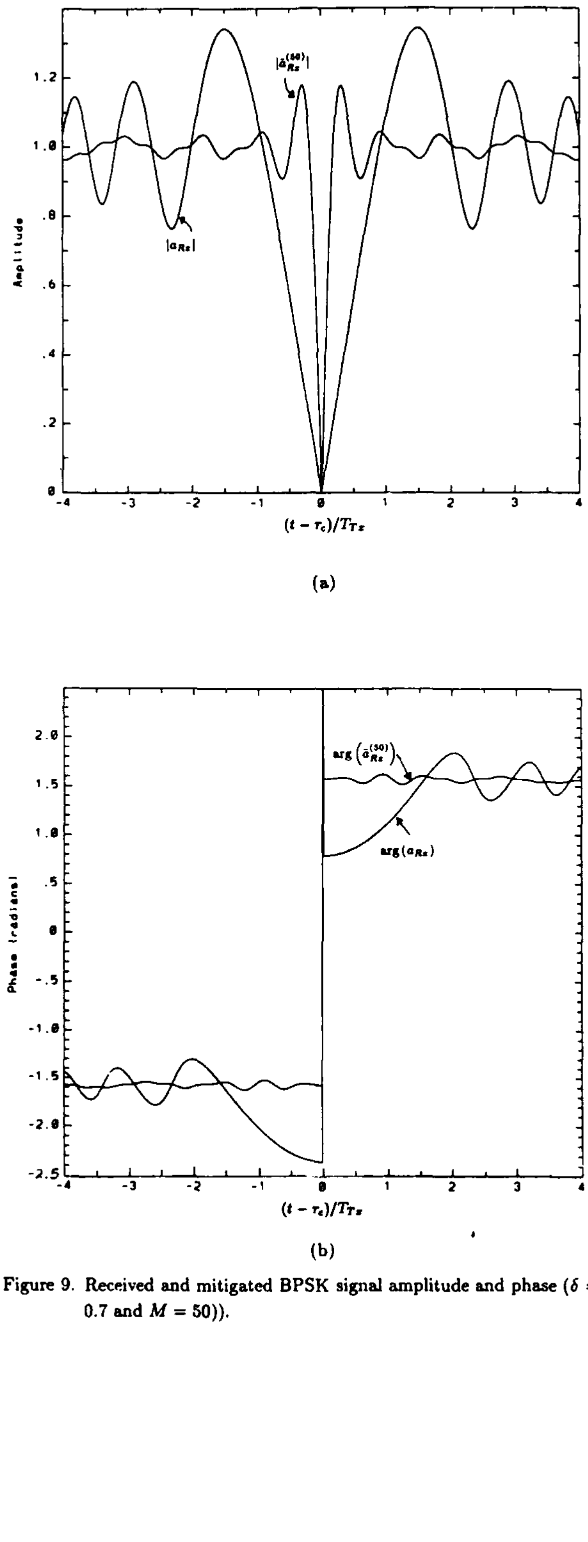


Figure 9. Received and mitigated BPSK signal amplitude and phase ($\delta = 0.7$ and $M = 50$)).

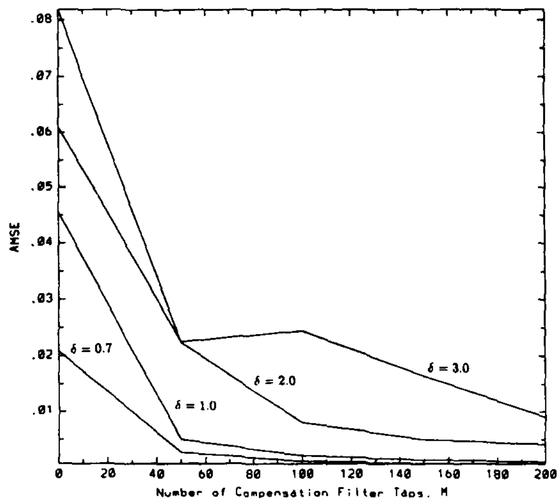


Figure 10. Pulse width error versus filter size for various dispersion factors.

5. ACKNOWLEDGMENTS

The authors wish to thank Prof. B. Dong for his contributions to this paper. This research work was sponsored by the U. S. Army Center for Command, Control and Communication Systems under contract DAAB-07-87K-A0006.

6. REFERENCES

[1] L. S. Wagner and J. A. Goldstein, "High resolution probing of the HF ionospheric sky channel: F2 layer results." *Radio Sci.*,vol **20**, pp 287–302, 1985.

[2] R. P. Basler, G. H. Price, R. T. Tsunoda and T. L. Wong,"HF channel probe." Proc. of 1987 Symposium on the Effect of the Ionosphere on Communication, Navigation, and Surveillance Systems, 5–7 May 1987.

[3] L. S. Wagner, J. A. Goldstein and W. D. Meyers, "Wideband probing of the transauroral HF channel: Solar minimum." Proc. of 1987 Symposium on The Effect of the Ionosphere on Communication, Navigation, and Surveillance Systems, 5–7 May 1987.

[4] C. C. Watterson, J. R. Juroshek and W. D. Bensema, "Experimental confirmation of an HF channel model." *IEEE Trans. Commun. Tech.*, vol **COM–18**, pp 792–803, 1970.

[5] L. Ehrman, L. B. Bates, J. F. Eschle and J. M. Kates, "Real-time software simulation of the HF radio channel." *IEEE Trans. Comm.*, vol **COM–30**, pp 1809–1817, 1982.

[6] M. Nessenbergs, "Modelling of wideband HF channels." Proc. of the 40-th Meeting of the Electromagnetic Wave Propagation Panel on Scattering and Propagation in Random Media, AGARD, Rome, Italy, 18-22 May 1987.

[7] D. G. Anderson and J. I. H. Askne, "Wave packets in strongly dispersive media." *Proc. IEEE*, vol **62** (11), pp 1518–1523, Nov. 1974.

[8] G. I. Terina, "On distortion of pulses in ionospheric plasma." *Radio Eng. and Electron. Phys.* (English translation), vol **12**, pp 109–113, 1967.

[9] Romanova and T. Ye. Ryzhkina, "Calculation of dispersive coefficients of pulsed signals for the oblique propagation of radio waves in the ionosphere." *Radio Eng. and Electron. Phys.* (English translation), vol **21**, pp 1–7, 1976.

DISCUSSION

C. GOUTELARD, FR
English Translation

You propose the use of a transverse filter to satisfy your matching requirements. Since we know that because of multi-path phenomena the ionospheric channel transfer function includes nulls, the function you are required to produce will have poles occuring in optimum fashion in a recursive structure. The transverse-only solution increases the number of factors - you use 100 - which in turn considerably increases the focusing time of the matching filter. Did you examine the possiblity of using a recursive structure or a combined structure? If so, why did you choose the parallel structure?

AUTHOR'S REPLY

We have not yet considered the approach of recursive filtering. As more complicated channels are considered in our analysis, this implementation may, indeed, become necessary in order to achieve fast adaptability to ionospheric changes.

IMPROVED RELIABILITY PREDICTION FOR HF DIGITAL VOICE TRANSMISSION

by
Dr. Alfonso Malaga
SIGNATRON, Inc.
110 Hartwell Avenue
Lexington, MA 02173-3198
USA

SUMMARY

A method for predicting the performance of High Frequency (HF) digital voice transmission is described. The method takes into account the effects of multipath propagation and signal processing techniques which exploit the multipath to achieve improved performance. A propagation prediction model which provides the necessary multipath and signal-to-noise ratio statistics is also described.

1. INTRODUCTION

Typical HF propagation models such as IONCAP and its predecessor HFMUFES predict reliable communications at those freqeuncies at which the signal-to-noise ratio (SNR) exceeds a desired level with high probability, say 90 percent. Often an additional requirement is that multipath propagation at these frequencies be minimal in order to avoid frequency selective fading. Consequently, these propagation models predict more reliable communications at frequencies which are close to the maximum useable frequency (MUF). However, with the advent of digital communications and adaptive signal processing, digital voice transmission systems, which exploit the diversity available due to multipath propagation, have been designed. These systems, if properly designed, often operate more reliably at those frequencies at which there is more multipath rather than at those frequencies at which the SNR is largest.

This paper proposes that reliable communications for HF digital voice transmission be predicted to occur when the probability that the bit error rate is smaller than, say, 10^{-3}, is greater than, say, 99%, with a confidence of 90 percent. Then the system is said to have 99% availability with 90% reliability.

Since the bit error rate of digital voice transmission at, say, 2400 bits per second, depends on the signal-to-noise ratio, the number of multipath components and their relative delays, and the modulation and signal processing employed, the paper describes a model for predicting these paramters as well at the system availability and reliability. The calculation of the system availability or equivalently, its outage probability for different types of modulation and signal processing which results in improved performance in the presence of multipath is discussed.

2. SYSTEM AVAILABILITY AND OUTAGE PROBABILITY

The propagation parameters which affect the performance of HF communications systems are the path loss and the multipath characteristics. A major objective of this paper is to define a system performance measure for HF digital systems which takes both of these into account. The communications user wants to know the fraction of the time that the system operates as desired (system availability, A). A digital system is said to be available when the bit error rate, P_b, is smaller than a threshold, b_c, say $b_c = 10^{-3}$. The bit error rate is a function of the signal-to-noise plus multipath interference ratio, ρ. Thus the system availability is defined as

$$A = \text{Prob}\{P_b < b_c\} = \text{Prob}\{\rho > \rho_c\} \qquad (1)$$

where ρ_c is the signal-to-noise plus multipath interference ratio which must be exceeded to achieve a bit error rate smaller than b_c. The outage probability, P_0 is defined as the fraction of the time that the system fails to operate reliability. Hence $P_0 = 1 - A$.

The parameter ρ is a random variable whose statistics depend on the HF channel statistics and the signal processing employed by the receiver. We will illustrate simplified methods for the calculation of ρ and A for two systems. One is a conventional digital modem which does not try to exploit the multipath diversity in the HF channel and the other is digital modem which uses adaptive equalization to combine the multipath coherently.

Conventional Modem Performance:

An example of a modem which does not try to exploit the benefits of multipath propagation is a modem which uses multiple QPSK modulated carriers (parallel tones) within the 3 kHz bandwidth. The signaling interval (pulse duration) of each carrier is much longer than the multipath spread of the channel. Time gating of the pulses is employed to avoid multipath interference between successive pulses as illustrated in

Figure 1. The main effect of the HF channel multipath is to produce Rayleigh fading. The fading across the entire 3 kHz band is correlated (i.e. all carriers fade simultaneously) if the multipath spread is small than 300 µs, as is the case when the operating frequency is near the MUF. When the multipath spread exceeds the time gating, there is intersymbol interference (ISI). Thus modems of this type try to operate near the MUF.

As a result of the multipath, the received signal has complex Gaussian statistics [1], and the detection SNR, ρ, is exponentially distributed with mean SNR given by

$$\lambda = \frac{\bar{S}}{N_0 R_b + \bar{I}} \tag{2}$$

where $\bar{S}$ is the average received signal power, N_0 is the average atmospheric plus man-made noise power density, R_b is the transmission bit rate and $\bar{I}$ is the average multipath interference power due to multipath components which exceed the time-gating protection in the system. The system availability is then

$$A_1 = \int_{\rho_c}^{\infty} \frac{1}{\lambda} e^{-\rho/\lambda} d\rho = e^{-\rho_c/\lambda} \tag{3}$$

and the outage probability $P_0 = 1 - A_1$.

Multipath Combining Modem Performance:

An example of a digital modem which uses multipath to improve system performance is illustrated in Figure 2. The modem employs a single QPSK modulated carrier. Therefore, the signaling interval (pulse duration) is much smaller than the multipath spread of the channel which results in ISI. The receiver employs an adaptive equalizer which removes the ISI and combines the multipath components so as to maximize the SNR plus ISI ratio. A mathematical model for predicting the performance of modems of this type for multipath fading channels has been derived by Monsen [2]. The detection SNR at the output of the adaptive equalizer can be expressed as

$$\rho = \sum_{i=1}^{M} \alpha_i \tag{4}$$

where M is the number of taps in the adaptive equalizer. The number of taps is typically chosen so that the delay spanned by the taps exceeds the maximum expected delay spread of the channel multipath.

When the channel bandwidth is 3 kHz, each resolvable multipath component (i.e. multipath components whose delay difference is smaller than the tap spacing) consists of an ordinary plus an extraordinary ray as well as micro multipath due to irregularities. Each resolvable multipath component has complex Gaussian statistics [1], and ρ is the sum of M uncorrelated random variables with exponential statistics [2]. The probability density of ρ can be written as [2]

$$p(\rho) = \sum_{i=1}^{M} \frac{b_i}{\lambda_i} e^{-\rho/\lambda_i} \tag{5}$$

where the λ_i represent the average signal-to-noise plus residual ISI ratio due to each resolvable multipath component which satisfy

$$\sum_{i=1}^{M} \lambda_i \leq \frac{\bar{S}}{N_0 R_b} . \tag{6}$$

The calculation of the λ_i is detailed in [2] and is a function of the transmit and receive filter impulse responses and the impulse response of the HF channel. In fact, Monsen [2] has shown that the λ_i are the eigenvalues of the MxM SNR matrix

$$\Lambda = \frac{\bar{S}}{N_0 R_b} G^{-1} C \tag{7}$$

where C is the signal covariance matrix whose diagonal elements represent the signal power (variance) present in each of the equalizer taps and the off-diagonal elements are cross-correlations, i.e.

$$C_{k\ell}(t_0) = \int_{-\infty}^{\infty} f(t_0 - kT_s - \tau) f(t_0 - \ell T_s - \tau) Q(\tau) d\tau, \quad k,\ell = 1,\dots,M \tag{8}$$

where f(t) is the impulse response of the cascade of the transmit and receive filters, t_0 is the sampling time, T_s is the equalizer tap spacing and $Q(\tau)$ is the mean power impulse response of the HF channel defined in terms of the normalized impulse response h(t) as

$$\overline{S}\cdot\overline{h(t)h^*(\tau)} = \delta(t-\tau)Q(\tau) \ . \tag{9}$$

The impulse response of the HF channel depends on the operating frequency and in general is of the form

$$h(t) = \sum_{i=1}^{L} a_i\delta(t-\tau_i) \tag{10}$$

where L is the number of multipath components, τ_i is the delay of i^{th} multipath component, a_i is its amplitude, and

$$\sum_{i=1}^{L} \overline{a_i^2} = 1.$$

The elements in matrix G in Equation (7) have two components, one which accounts for the background noise and the other accounts for the intersymbol interference from previous and future transmitted symbols, i.e.

$$G_{k\ell} = N_{k\ell} + \frac{\bar{S}}{2N_0R_b} \sum_{j\neq 0} C_{k\ell}(t_0-jT) \tag{11}$$

with

$$N_{k\ell} = \int_{-\infty}^{\infty} f_R(kT_s-\ell T_s-t)\ f_R(t)\ dt, \quad k,\ell=1,\ldots,M \tag{12}$$

where $f_R(t)$ is the impulse response of the receiver filter, T_s is equalizer tap spacing, and T is the transmitted symbol duration.

The system availability is found from

$$A_2 = \int_{\rho_c}^{\infty} p(\rho)d\rho = \sum_{i=1}^{M} b_i e^{-\rho_c/\lambda_i} \tag{13}$$

where the λ_i are the eigenvalues of Equation (7) and the b_i are found by applying partial fraction expansion techniques or residue theory to the equality:

$$\prod_{i=1}^{M} (1 + \lambda_i s)^{-1} = \sum_{i=1}^{M} b_i(1 + \lambda_i s)^{-1} \tag{14}$$

The system availability expression in Equation (13) indicates that the modem performance improves at the number of significant eigenvalues λ_i increases; that is, it improves as the number of resolvable multipath components increases.

3.0 PREDICTION OF CHANNEL MULTIPATH PARAMETERS

In order to predict the availability of multipath combining modems we need to determine the number of multipath components, their amplitudes and relative delays, and we need to predict the noise level. Prediction of the number of multipath components requires knowledge of the critical frequencies of the E, F1,and F2 layers and their heights. Some models for prediction of these parameters, e.g. IONCAP [3], rely on world-wide maps of the critical frequencies generated from measurements at a few locations during periods of low and high sunspot activity. Other models, e.g. MINIMUF [4] and extensions of it [5], rely on theoretical models of the diurnal and seasonal behavior of the critical frequencies and heights of the layers coupled with measurements to determine their sunspot *dependence*. *Given an* acceptable model of the ionosphere, we can predict the multipath structure as follows.

Skywave Path Take-Off Angles:

Neglecting the effects of the Earth's magnetic field, we first determine the elevation take-off angles Δ_i of each multipath component by solving the equation

$$D_k = 2R_e \sin\frac{D}{2R_eK} = 2\tan\theta_i \int_0^{h_r} [1 - \{\frac{f_p(h)}{\cos\theta_i f}\}^2]dh \tag{15}$$

where D is the great circle distance between transmitter and receiver, D_k is a horizontal (flat-earth equivalent) range as shown in Figure 3, θ_i is the angle of incidence on the ionosphere, K is the number of hops, R_e is the radius of the earth, f is the operating frequency, h_r is the height of reflection of the ray measured relative to the flat-earth range D_k, and $f_p(h)$ is the ionospheric plasma frequency height profile (see Figure 4) which varies with time-of-day, season, etc.

For a fixed path length D, operating frequency f, time-of-day, season, etc., there are a number of angles (θ_i, Δ_i) which satisfy Equation (15). Each solution corresonds to a skywave path (ray). The expression for calculation of the path take-off angles ignores the effects of the earth's magnetic field which is a good approximation for the ordinary rays. The expression for the extraordinary rays is more complicated [6].

Skywave Path Delays:

When the earth's magnetic field is neglected, the delay for each of the skywave paths can be determined from [6]

$$\tau_i(\theta_i) = \frac{KD_K}{c\ \sin\theta_i} \tag{16}$$

where D_K is the horizontal flat-earth equivalent range defined in Equation (15), K is the number of hops, θ_i is the angle of incidence of the ray on the ionosphere, and c is the speed of light (3×10^8 m/sec).

Skywave Path Amplitude

The amplitude of each skywave return depends on the propagation losses, $L(\theta_i)$ and a number of equipment related factors such as transmitter power, P_t, transmit antenna radiation efficiency, η_1, transmit antenna impedance match efficiency, η_2, and the directivity patterns of the transmitting and receiving antennas, $G_t(\Delta_i)$ and $G_r(\Delta_i)$ respectively. The efficiency of the receiving antenna also affects the amplitude of each ray. However, since atmospheric and man-made noise are the primary sources of noise at HF and these are affected equally by the receiving antenna efficiency, the latter can be ignored in estimating the signal-to-noise ratio (SNR) of each skywave return. Thus,

$$10\ \log_{10}(\overline{a_i^2 S}) = P_t + \eta_1 + \eta_2 + G_t(\Delta_i) + G_r(\Delta_i) - L(\theta_i) \tag{17}$$

where all quantities are in dB.

The radiation efficiency of the antenna is frequency dependent and difficult to quantify, especially when the antenna is electrically small (smaller than an eight of a wavelength), as its radiation and ohmic loss resistance are affected by the ground conductivity below the antenna.

The propagation losses of each skywave return are due to four effects: 1) the spherical spreading loss L_S; 2) the absorption loss L_A; 3) the ground reflection loss L_G when the ray is a multi-hop ray; and 4) the polarization mismatch loss L_P.

The spherical spreading loss is slightly smaller than that of a wave propagating in free space because of the focusing effect associated with reflection from an inhomogeneous medium. It can be expressed in dB as

$$L_S = 32.45 + 20\ \log f + 10\ \log(KD_K/\sin\theta_i) + 10\ \log\{K\cos\theta_i\left|\frac{dD_K}{d\theta_i}\right|\} \tag{18}$$

where the frequency f is in MHz, and D_K is in km. The third term in Equation (18) is the spreading loss in the azimuth plane while the fourth term is the spreading loss in the elevation plane. This last term is not valid near the caustic, i.e., at frequencies near the MUF of the layer where the ray is reflected. At these frequencies a correction term can be used [6].

The spreading loss is the main source of attenuation on HF paths, especially at night. However during the day time and at high latitudes at all times of day, absorption in the D-region and E-region of the ionosphere can contribute significantly to the path loss. The absorption loss in dB can be expressed as

$$L_A = \frac{K\ G(\chi,R,Y)\ \sec\theta_i}{(f \pm f_L)^2 + (\nu/2\pi)^2}[\alpha_D + \alpha_E B(\cos\theta_i f/f_0E)] \tag{19}$$

where the upper sign is for ordinary rays and the lower for extraordinary rays. The first term in brackets accounts for non-deviative absorption in the D-region and the second term accounts for deviative absorption in the E-region. The factor $G(\chi,R,Y)$ describes the dependence of the absorption loss on the solar zenith angle, χ, the sunspot number, R, and geomagnetic latitude Y. These variations are due to variations in the electron-neutral collision frequency and the critical frequency of the E-layer. The factor $B(\cdot)$ accounts for ray bending of the wave in the E-layer, and is proportional to the difference between the group path $P'(\cos\theta_i f/f_0E)$ and phase path $P(\cos\theta_i f/f_0E)$ due to the propagation through the E-layer, i.e. the difference in group and phase paths after propagation through the F1 and F2 layers is neglected because the collision frequency at those heights is negligable. Hence

$$B(x) = [(\frac{\nu \cos\theta_i}{2\pi f_0 E})^2 + (x \pm \frac{f_L \cos\theta_i}{f_0 E})^2][P'(x) - P(x)] \quad (20)$$

with $f_L = f_H \sin I$, where f_H is the gyrofrequency, I is the magnetic dip angle, f_0E is the E-layer critical frequency, and

$$P'(x) - P(x) = \begin{cases} \frac{1}{4x}(1 + x^2) \ln\frac{1 + x}{1 - x} - \frac{1}{2}, & \text{if } x < 1 \text{ (E-layer reflection)} \\ \frac{1}{2x}(1 + x^2) \ln\frac{1 + x}{x - 1} - 1, & \text{if } x > 1 \text{ (F1 and F2 layer reflection)} \end{cases} \quad (21)$$

The parameters G, ν, α_D, and α_E completely specify the absorption loss. The prediction model HFMUFES [7] assumes that the absorption in the D and E regions is non-deviative, i.e. it assumes $B(x) = 1$, $\alpha_D + \alpha_E = 1$, and $(\nu/2\pi)^2 = 10.3$, while $G(\chi,R,y)$ is based on empirical data. The new prediction model, IONCAP [3], uses a similar model for F layer low-ray reflections except that $G(\chi,R,y)$ has been replaced by a different model which relates its solar zenith and sunspot dependence to that of f_0E. IONCAP uses correction factors to account for additional losses for the E-layer reflections and high-ray F-layer reflections. These correction factors presumably account for additional deviative absorption losses. A third method developed at Appleton Laboratories [8] uses an expression similar to Equation [19] with $\nu = 0$, and with $G(\chi,R,y)$ based on empirical data.

The ground reflection loss for multi-hop rays depends on the polarization of the rays. Both ordinary and extraordinary rays are in general elliptically polarized with their polarization rotating in opposite senses. Since an arbitrary elliptically polarized wave can be decomposed as the sum of a vertically and a horizontally polarized component, the ground reflection loss for a K-hop ray can be expressed in dB as

$$L_G = -(K - 1)\ 10 \log\{\frac{|R_V(\Delta_i)|^2 + M_W^2|R_H(\Delta_i)|^2}{1 + M_W^2}\} \quad (22)$$

where M_W is the magnitude of the ray polarization vector (ratio of polarization ellipse axes), R_V is the reflection coefficient for the vertical component of the downcoming wave, and R_H is the reflection coefficient for the horizontally polarized component. When the downcoming ray is linearly polarized, $M_W = 0$ if the wave is vertically polarized and $M_W = \infty$ if it is horizontally polarized. When the downcoming wave is circularly polarized $M_W = 1$.

The polarization mismatch loss occurs because the polarization of the transmitting and receiving antennas is rarely matched to the characteristic polarization of a ray. Furthermore, the polarization of a wave changes as it propagates through the ionosphere so that even if the transmitting antenna is matched to the characteristic polarization of the ray at its point of incidence on the ionosphere, then a receiving antenna of the same polarization will not be matched to the polarization of the ray as it emerges from the ionosphere. The polarization mismatch loss is given in dB by

$$L_p = -10 \log\{\frac{(1+M_A M_W)^2\cos^2\psi+(M_A+M_W)^2\sin^2\psi}{(1+M_A^2)(1+M_W^2)}\} \quad (23)$$

where M_A is the magnitude of the antenna polarization vector and M_W is the magnitude of the characteristic polarization of the ray at its point of incidence on the ionosphere if the antenna is the transmitting antenna, and at its point of exit if the antenna is the receiving antenna. The polarization of a wave depends on its direction of propagation relative to the earth's magnetic field and the frequency. It is given in [6].

Background Noise Levels

Background noise at HF can be classified as man-made, and atmospheric noise. IONCAP [3] estimates median values of man-made noise according to the expression

$$N_M = N_0 + 29 \log_{10}(f/3) \quad (24)$$

where N_M is the man-made noise power in a 1 Hz bandwidth, f is the frequency in MHz, and N_0 is the measured man-made noise power at 3 MHz (in a 1 Hz bandwidth). Typical values of N_0 in remote, rural, residential, and industrial are -163 dBW/Hz, -148 dBW/Hz, -135 dBW/Hz, and -125 dBW/Hz, respectively. These values are about 10 dB higher than those used in HFMUFES [7]. The upper and lower decile variability values of man-made noise are taken as 9 dB and 7 dB from the median.

Atmospheric noise levels can be obtained from the worldwide 1 MHz noise maps, frequency dependencies and variability charts found in CCIR Report 332 [9].

4.0 AVAILABILITY AND OUTAGE PROBABILITY PREDICTION

The propagation model of Section 3 has been used to calculate the outage probability for the two types of modems described in Section 2 assuming uncoded QPSK transmission at 2400 bps and a bit error rate outage threshold $b_c = 10^{-3}$. The predicted outage probability over a 1600 km mid-latitude path is plotted in Figures 5 as a function of the operating frequency. The predictions assume 400 W effective radiated power at all frequencies, quarter-wave monopole antennas above ground of average conductivity, sunspot number equal to 100, mid-March propagation conditions at 1900 hrs, and median signal-to-noise ratios. The improvement in system availability (1-outage probability) when the modem combines the multipath coherently is evident from the charts. The charts show that the optimum performance when no multipath combining is employed is achieved by operating at frequencies near the MUF where the SNR is largest. Multipath combining modems outperform conventional modems at frequencies above the FOT (the frequency of optimum transmission where there is no multipath, i.e. 11-13 MHz) because of the diversity afforded by F2 high-ray/low-ray multipath, and at frequencies below the FOT because of the diversity from 1 and 2 hop F2 ray multipath. Figures 6, 7, and 8 show the multipath characteristics of the HF channel at the frequencies where the multipath combining modem achieves optimum performance (10 and 15 MHz) and at 11 MHz which is the frequency that prediction models such as IONCAP and HFMUFES refer to as the FOT.

From Figure 6 we can also see that if we assume that the required system availability is 99% (i.e. $P_0 = 10^{-2}$), then the lowest useable frequency (LUF) at 1900 hours for the multipath combining modem is 7.5 MHz versus 8.5 MHz for the conventional modem. This shows that the ability to combine multipath coherently extends the range of useable frequencies.

5.0 RELIABILITY

The availability predictions presented in this paper assume median SNRs. Therefore, the system will achieve the predicted availability at the selected time, season, etc. with 50% confidence. Higher confidence in the prediction requires knowledge of the distribution of the signal-to-noise ratio and multipath environment as a function of frequency, path length, geographic location, time-of-day, season, etc. Distributions of the atmospheric noise level as a function of time-of-day, season, and geographic location can be found in [9]. Similar data for the signal strength is not available. However, if the signal strength variability is due to variability in ionospheric absorption, then periods in which the signal level decreases will result also in lower atmostpheric noise levels and man-made noise levels if the latter propagate via skywaves also. This implies that use of the atmospheric noise distributions to approximate the SNR distribution would result in overly pessimistic results. The same would be true if we were to use path loss distributions as done in HFMUFES [7]. A true SNR distribution model is needed in order to improve the reliability of HF system performance predictions.

REFERENCES

[1] C. Watterson, et al., "Experimental Confirmation of an HF Channel Model", IEEE Trans. Comm. Tech., Vol. V-COM-18, December 1970, pp. 792-803.

[2] P. Monsen, "Theoretical and Measured Performance of DFE Modem on a Fading Multipath Channel", IEEE Trans. Comm., Vol. V-COM-25, October 1977, pp. 1144-1153.

[3] J.L. Lloyd, et al., Institute for Telecommunication Sciences, "Estimating the Performance of Telecommunication System Using the Ionosphere Transmission Channel", 1978, NTIA Report 83-127.

[4] P.H. Levine, et al., "A Simplified HF MUF Predicition Algorithm", Proc. IEE Conf. Antennas and Prop., Pub. No. 169, Pt. 2, November 1978, pp. 161-167.

[5] A. Malaga, "A Global Model for Wideband HF Skywave Propagation", Proc. 1981 Symp. on the Effect of the Ionosphere on Radiowave Systems, April 1981, pp. 553-364.

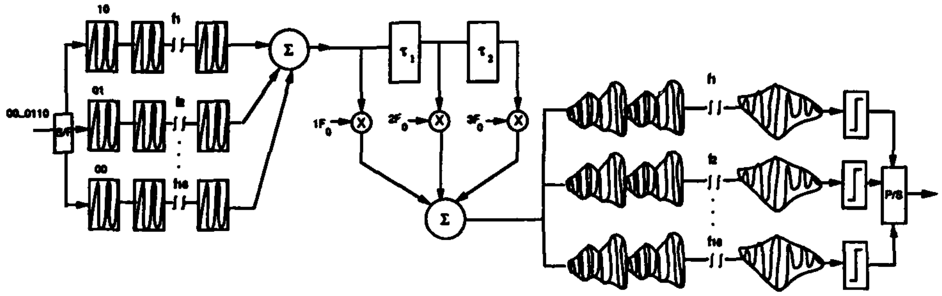
[6] K.G. Budden, Radio Waves in the Ionosphere, Cambridge, UK, University Press, 1961.

[7] A. Barghausen, et al., Institute for Telecommunication Sciences, "Predicting Long Term Operational Parameters of HF Skywave Telecommunication Systems", May 1969, Technical Report ERL-ITS-78.

[8] P.L. George and P.A. Bradley, "A New Method of Predicting Ionospheric Absorption", Telecommunication Journal, Vol. V-41, 1974, pp. 308-312.

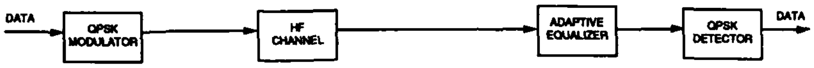
[9] CCIR Report 322, "World Distribution and Characteristics of Atmospheric Radio Noise", International Radio Consultative Committee, Documents of the Xth Plenary Assembly, Geneva, 1963, ITU, Geneva, 1964.

(a)

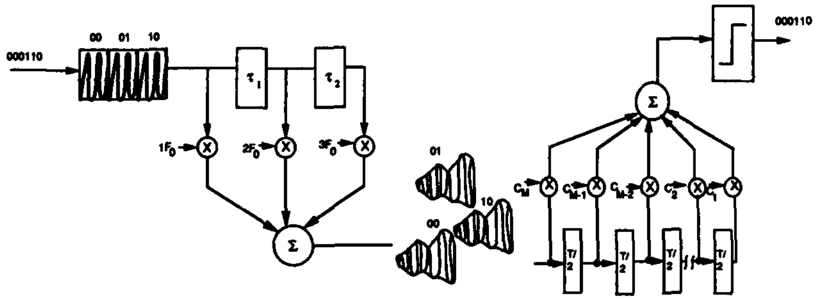


(b)

Figure 1 Illustration of Parallel Tone Modem Processing with Time Gating for HF Channel Multipath Protection

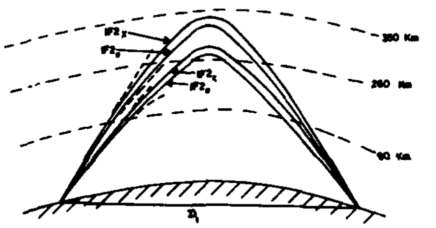


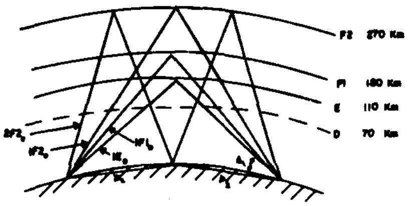
(a)



(b)

Figure 2 Illustration of Adaptive Equalizer Modem Processing to Remove ISI Induced by HF Channel Multipath





(a) (b)

Figure 3 HF Multipath from Multiple Reflections from a Single Layer (a) and Multi-Hop Reflections from Multiple Layers (b)

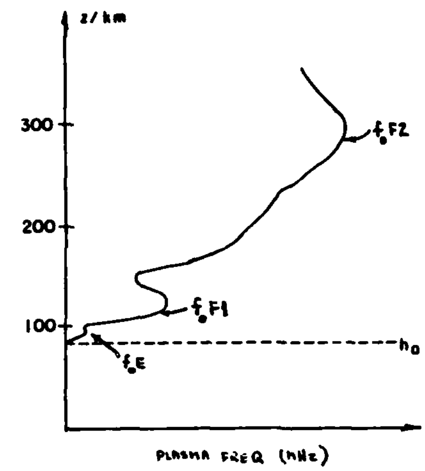


Figure 4 Typical Ionospheric Plasma Frequency Height Profile

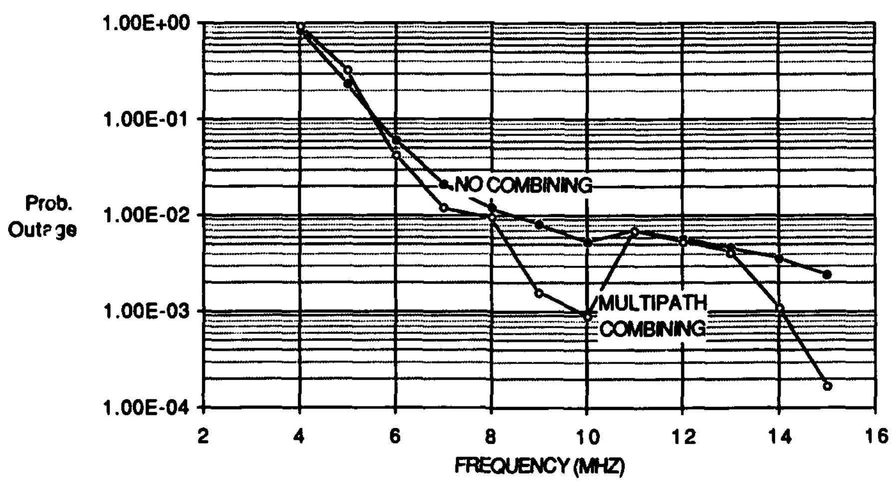


Figure 5 Outage Probability vs Operating Frequency for 1600 km Link at 1900 hrs During Mid-March (SSN = 100), Assuming 400 W Effective Radiated Power

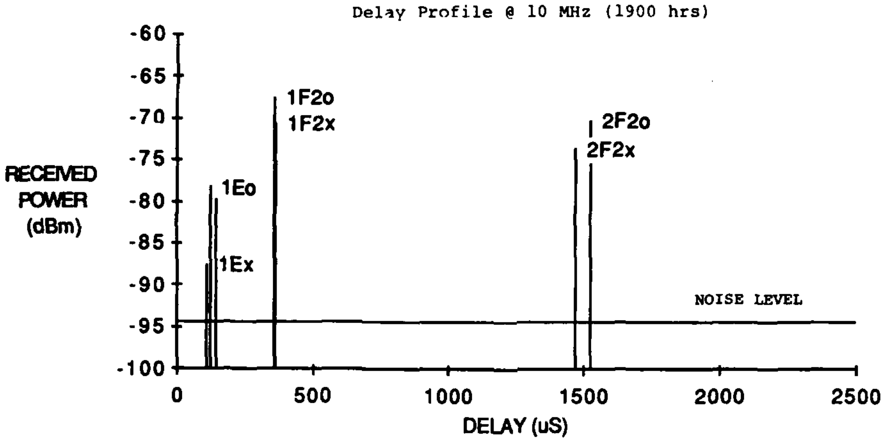


Figure 6 Impulse Response of 1600 km Link and Background Noise Level at 10 MHz

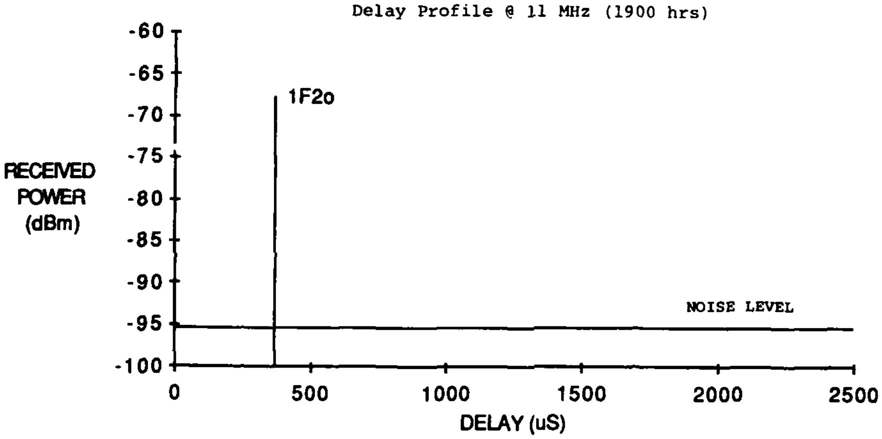


Figure 7 Impulse Response of 1600 km Link and Background Noise at 11 MHz

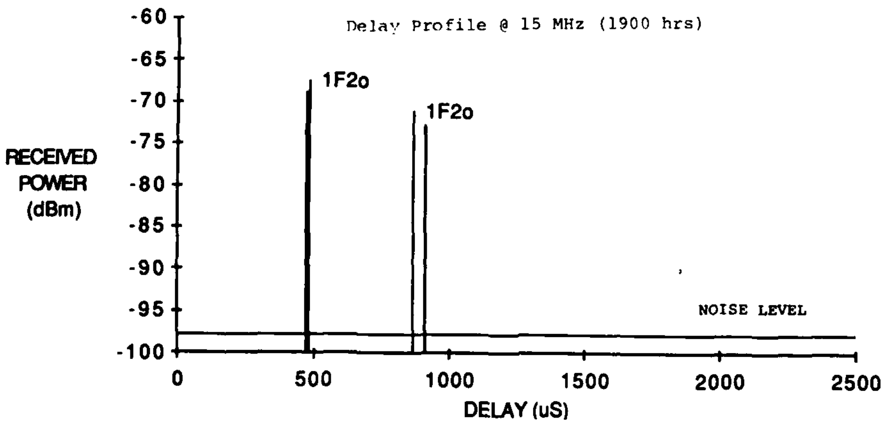


Figure 8 Impulse Response of 1600 km Link and Background Noise at 15 MHz

DISCUSSION

J. BELROSE, CAN

My comment relates not only to your paper, but to the previous paper also (Yeh and Mickelson). You at least made an attempt to predict the real world, but not in sufficient detail to be useful. The ionospheric channel is a time variable frequency selective and frequency spread channel. Noise and interference are important difficulties with HF communications. The ionospheric model that you describe is an average model which is not very useful to assess the ability of the channel to transmit digital voice. Another problem relevant to HF digital voice concerns the available bandwidth. In order to employ a rate of 2400 bits/second, it is necessary to employ some form of linear predictive coding (LPC) technique, and such systems do not degrade gracefully as signal-to-noise decreases, digital error rate increases. You did not touch on this topic.

Concerning HP channel simulations - I call your attention to my paper presented at our 1988 Fall Meeting in Paris.

AUTHOR'S REPLY

All HF performance prediction models currently in use are "average" statistical models and none of them consider the frequency selective effects due to multipath. The main point of the paper is to define a performance measure which accounts for frequency selective fading effects. The only assumption made regarding the multipath combining modem performance is that the modem will be able to track the time selective fading due to frequency spreading. The inability of certain model implementations to track the fading can only be determined using channel simulators. The performance measure defined is general and applies to linear predictive digital voice encoding as well as any other applicable technique. All one needs to know is the bit error rate required by the encoding technique. If the bit error rate exceeds the maximum bit error rate that LPC (or any other technique) can tolerate, the system is assumed to suffer an outage; that is, no graceful degradation is assumed.

G. HAGN, US

You stated that there is no model for HF interference, but this is only partially correct. The HF other-user interference, specified as congestion of allocated subbands, has been successfully modeled by Profs. P. Laycock and G. Gott at UMIST, UK (see the AGARD Conference Proceedings, Lisbon, 1987). The UMIST model features the derivation of coefficients from empirical 1-kHz channel occupancy data for a given subband. An attempt is now being made to obtain occupancy and congestion data in the UK and at other sites in Western Europe (e.g. Sweden) in order to develop a regional model of congestion. The interpolation and extrapolation of the coefficients to locations where no data had been taken would form the basis of such a regional model.

AUTHOR'S REPLY

If this work is successful, then there would be a usable model for HF interference in Western Europe. The concept could be generalized to form a global model based on a collection of such regional models, but much empirical data would be required.

A MODERN TRANS-IONOSPHERIC PROPAGATION SENSING SYSTEM

G. J. Bishop and J. A. Klobuchar
Ionospheric Physics Division
Geophysics Laboratory, Air Force Systems Command
Hanscom Air Force Base, MA 01731

Lt Col A. E. Ronn and Capt M. G. Bedard
Headquarters Air Weather Service
Scott AFB. IL 62225-5008
USA

SUMMARY

One of the most important potential problems with modern military systems which utilize spacecraft in various ways is the effect of the ionosphere on the radio signals which pass to and from the spacecraft. Such systems include active communications and navigation satellites as well as both ground-based and potential space-based ranging systems. The major effects the ionosphere can have on such systems are the additional time delay the electrons in the earth's ionosphere add to the free space path delay, the short term rate of change of this additional delay, amplitude scintillation or fading effects the signal encounters due to irregularities in the ionosphere, and Faraday rotation of linearly polarized radio waves transmitted through the ionosphere. While some of these effects have been studied since the beginning of the artificial satellite era in the late 1950s, adequate models of these effects on military systems still do not exist. This lack of suitable models is due in part to the high temporal and geographic variability of the ionosphere, and partly due to the increasing radio signal accuracy requirements and special geometries of various military systems which must propagate through the ionosphere.

The US Air Force's Air Weather Service is procuring a modern trans-ionospheric sensing system, called "TISS", which will consist of a number of stations located throughout the world, making real time measurements of the time delay of the ionosphere, and its rate of change, as well as amplitude scintillation, along several different viewing directions from each station. These trans-ionospheric measurements will be used to allow models, which currently provide only monthly climatology of these parameters, to provide real time specification of trans-ionospheric propagation parameters. The real-time specifications of these parameters can then be used as decision aids in both the tactical and the strategic military environments. The TISS will include first order artificial intelligence design to aid in gathering the most appropriate sets of available real-time trans-ionospheric propagation data, and will communicate these data sets to the Air Weather Service's Space Forecasting Center where they will be tailored to specific military customers.

I. TRANS-IONOSPHERIC EFFECTS ON RADIO WAVES

There are several potential effects which the ionosphere can have on radio waves which must propagate through it. Modern military systems can be limited in performance by one, or more, of these effects and systems planners need to take their potential impacts into account, both in system operation and in new system design. Ideally, system designs will contain "work-arounds", such as alternate channels, additional space vehicles, or robust signal formats, that mitigate impacts of ionospheric effects on system performance. However, since economic and other constraints may not allow mitigation in the system design, system operators require near-real-time specification of ionospheric effects. With this input, operators can determine system status and can employ operational mitigation, such as longer integration times, alternate look directions, or reliance on alternate systems, or time periods.

The types of military systems that require consideration of trans-ionospheric effects include: communications, navigation, and ground or space-based surveillance systems, that depend upon radio-wave propagation, (see Table 1). Significant studies of ionospheric effects have been performed supporting satellite communications[1], single frequency navigation using the Global Positioning System, (GPS)[2], and surveillance[3].

TABLE 1. **EXAMPLES OF SYSTEMS WHICH MAY REQUIRE CONSIDERATION OF TRANS-IONOSPHERIC EFFECTS**

TRANS-IONOSPHERIC SYSTEM	SYSTEM FUNCTION
SATELLITE COMM. LINKS	COMMUNICATION
GROUND-BASED RADARS (GBR)	SURVEILLANCE
SPACE-BASED RADARS (SBR)	SURVEILLANCE
GLOBAL POSITIONING SYSTEM (GPS) (single frequency user only)	NAVIGATION

The major trans-ionospheric effects on radio signals at typical military system frequencies are: 1) delay of the signal, 2) Faraday rotation of linearly polarized signals, 3) rapid changes in signal delay, and 4) signal fading. These effects are caused by different ionospheric parameters. Specifically, 1) above is due to the total number of electrons along the signal path through the ionosphere, where the number of electrons is measured in an equivalent square meter column, and is defined as the Total Electron Content, or TEC, with 1×10^{16} electrons/m^2 defined as one TEC unit;
2) is due to the product of the total number of electrons multiplied by the longitudinal component of the earth's magnetic field along the radio signal path. Phase scintillation, 3), is due to the time rate of change of the electron content, and 4) is produced by diffraction from irregularities of scale size of approximately one kilometer, along the path. Depending upon the system design, these major effects have the potential to degrade system performance in varied ways, such as: data loss, ranging errors, and degradation of detection, tracking or imaging, see Table 2.

TABLE 2. **POTENTIAL IONOSPHERIC IMPACTS ON SYSTEMS**

SYSTEM FUNCTION	POTENTIAL IMPACT	IONOSPHERIC RF EFFECT
COMMUNICATION	DATA LOSS	SIGNAL FADING
NAVIGATION	POSITIONING ERRRORS	SIGNAL DELAY
SURVEILLANCE - GROUND-BASED	RANGE, ORBIT, OR TRAJECTORY ERRORS	SIGNAL DELAY
- SPACE-BASED	DEGRADED RETURNS OVER WIDE REGIONS	FARADAY ROTATION
- BOTH	DEGRADED DETECTION, TRACKING OR IMAGING	FADING, RAPID CHANGES IN SIGNAL DELAY

The magnitude of any potential impacts on system performance due to trans-ionospheric effects is highly dependent on the specific parameters of the particular system in question. For example, it is generally true that the higher the operating frequency the smaller the effect of the ionosphere. However, other considerations such as technology limitations, and frequency band availability, may prevent resolving all ionospheric effects questions by raising the operating frequency. Table 3 illustrates some specific systems parameters and mitigation approaches that may be applied. Some mitigation approaches, such as for dwell time, may conflict, making complete mitigation more difficult and near-real-time monitoring of ionospheric effects more valuable.

TABLE 3. **SOME SYSTEM PARAMETERS AND MITIGATION APPROACHES FOR IONOSPHERIC EFFECTS**

PARAMETER	MITIGATION APPROACH
FREQUENCY	RAISE FREQUENCY
POLARIZATION	CIRCULAR POLARIZATION
FADE MARGIN	INCREASE MARGIN
DWELL TIME	DECREASE DWELL (Phase Effects)
	INCREASE DWELL (Fading Effects)
RANGE MARGINS	INCREASE MARGINS
REDUNDANCY	ADD COMM. CHANNELS AND OBSERVATION PLATFORMS

To better understand this variability of the effects the ionosphere can have on trans-ionospheric propagated waves at VHF, or higher frequencies, a more detailed understanding of the individual effect is required.

1. <u>Signal Delay, Causing Range Errors</u> This effect on ranging systems is due to ionospheric group delay, which is proportional to the electron content of the ionosphere. The magnitude of this group delay range error is proportional to the inverse square of the system operating frequency. This error frequently can be over 4 kilometers at a frequency of 100 MHz, which corresponds to 40 meters at a system operating frequency of 1 GHz. Worst case errors can be a factor of at least two higher than this, and when traversing the ionosphere at a low elevation angle, this error is further multiplied by up to an additional factor of three due to increase of the signal

path length in the ionosphere. Thus, ionospheric group delay errors can be as large as 250 meters, even at a system operating frequency of 1 GHz.

2. Range-Rate Errors are due to the time rate of change of the electron content of the ionosphere as seen by the observing system. The range-rate error depends upon the diurnal rate of change of the electron content of the ionosphere, upon the structure of any large-scale irregularities that may exist in the region, and upon the motion of the line of sight through the ionosphere. For instance, a satellite moving up from the horizon will usually encounter fewer electrons as it rises in elevation, simply due to the decrease of the signal path length in the ionosphere. These geometric changes are usually much greater than any diurnal changes in the electron content of the ionosphere.

3. Faraday rotation of linearly polarized radio waves, which can cause signal attenuation due to cross polarization effects, is proportional to the electron content of the ionosphere times the longitudinal component of the earth's magnetic field. The magnitude of this effect is proportional to the inverse square of the system operating frequency. At a frequency of 100 MHz, many radians of Faraday rotation can occur, and even at 1.3 GHz, the two-way Faraday rotation can often exceed 90^{o} during times of high solar activity. When the rotation is near 90^{o} nearly complete cross polarization signal loss occurs.

4. Rapid Changes in Signal Delay or phase scintillations, are due largely to rapid changes in the electron content of the ionosphere. These changes can be due solely to temporal changes in electron content, or can be a combination of geometric changes due to a vehicle moving such that the ray path moves through a large gradient in electron content during the observing period.

5. Fading (and enhancements) or amplitude scintillation of the amplitude of the received signal, is due to kilometer size irregularities in the electron density of the ionosphere. This fading is characterized by a statistical description of the percentage of time below which the signal fades, and is described by Whitney[4] and Aarons[5].

There are other effects which the ionosphere can have upon the characteristics of radio waves which propagate through it, such as distortion of the modulation envelope and angular refraction, or bending, of radio waves. However, these two effects are generally only potential problems at low VHF and only for large percentage signal bandwidths, or low elevation angles, respectively.

Most of the potential effects listed above are proportional, in some manner, to the total number of electrons through which the radio wave travels on its way from its transmitter to the receiver. A more complete discussion of the magnitude of many of these potential effects is available in the Handbook of Geophysics and the Space Environment, Sections 10.7, 10.8 and 10.9[6]. Table 4 summarizes the major trans-ionospheric effects and their interaction with systems.

TABLE 4. **TRANS-IONOSPHERIC EFFECTS ON SYSTEMS**

IONOSPHERIC PARAMETER	RADIO (RF) EFFECTS	SYSTEMS EFFECTS	POTENTIAL DEGRADATION OF:
TOTAL ELECTRON CONTENT (TEC)	SIGNAL DELAY	RANGE ERROR	TARGET LOCATION
	FARADAY ROTATION	SIGNAL LOSS	SURVEILLANCE
AMPLITUDE SCINTILLATION	FADES & ENHANCEMENTS	SIGNAL FADES TARGET FADES	MESSAGE CONTENT TARGET UPDATE
PHASE SCINTILLATION	RAPID CHANGES IN SIGNAL DELAY	RAPID CHANGES IN APPARENT RANGE	DETECTION TRACKING IMAGING

II. MODELS OF IONOSPHERIC EFFECTS

The ionosphere exhibits high temporal and geographic variability. This variability may be categorized by geographic region and by temporal periods. These are listed in Table 5. Temporal variations range from variations within the visible ionosphere, which may drift past in minutes, or may appear as observation geometry changes, to variations over the approximately 11 year solar cycle. The geographic regions, (Figure 1), vary in their exact boundaries with prevailing geophysical conditions. The mid-latitude region exhibits the fewest disturbances. For example, scintillation effects are comparatively infrequent in this region, and the diurnal behavior of TEC can be more accurately projected statistically. The trough region, occurring at the northern boundary of the mid-latitude region, is a geographic area of varying width where the ionization is depleted dramatically from levels in surrounding regions. The auroral, and polar cap regions frequently are the host regions for major disturbances caused by particle precipitation, guided along magnetic field lines, from the sun. These disturbances can propagate down to the mid-latitudes as well. The equatorial region also exhibits major disturbances generated through different processes which have a strong diurnal

TABLE 5. **IONOSPHERIC EFFECTS AND SOURCES OF THEIR VARIATION**

VARIATION CATEGORY		MAJOR IONOSPHERIC EFFECTS
GEOGRAPHIC	TEMPORAL	SIGNAL DELAY
POLAR CAP	MINUTES (GEOMETRY)	FARADAY ROTATION
AURORAL ZONE	DIURNAL	RAPID CHANGES IN DELAY
TROUGH	SEASONAL	FADES AND ENHANCEMENTS
MID-LATITUDE	SOLAR CYCLE	
EQUATORIAL	MAGNETIC STORM	

dependence. The variations in these regions, and the mechanisms that cause them are subjects of continuing theoretical and experimental studies. Even the monthly average conditions of the auroral and polar cap regions are not yet fully modeled, or understood.

Trans-ionospheric propagation effects have been intensively studied since the beginning of the artificial satellite era in the late 1950's. While many models have been constructed of electron density profiles, or only of the density at the peak of the F2 region, the electron density model of choice used by the Air Weather Service is a relatively new model called the Ionospheric Conductivity and Electron Density, (ICED), model[7]. The ICED model is a statistical model of the large scale features of the ionosphere, which includes features for the separate physical processes which are known to exist in different regions of the ionosphere. There are different algorithms, for example, for low-latitudes, mid-latitudes, the sub-auroral trough, and the polar cap. The ICED model currently represents only median climatology, but is soon to be updated to include near-real-time TEC measurements from TISS, as well as near-real-time bottomside electron density profiles from the DISS.

For amplitude and phase scintillation the WBMOD model[8], (Secan, et. al., 1987) currently represents the best effort at state-of-the-art monthly average climatology. WBMOD is currently being improved to include the measured amplitude and phase scintillation measurements from the TISS.

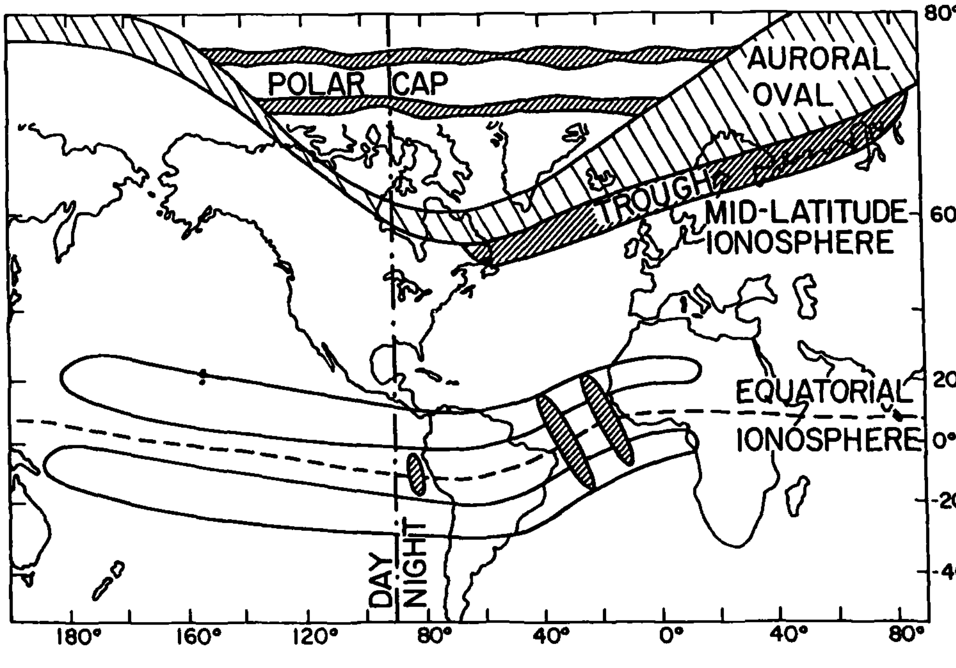


Figure 1. MAJOR GEOGRAPHIC REGIONS OF THE IONOSPHERE

The increasing radio signal accuracy requirements and special geometries of various modern military systems are drivers for these and further model upgrades. Such efforts strive not only to improve accuracy and timeliness of specifying ionospheric effects, but also to be able to better tailor environmental products to the requirements of specific systems. For example, a surveillance system may conduct the majority of its operation at low elevation angles, such as 10 degrees or less. Figure 2 illustrates this geometry, approximately to scale, for ground-based radars (GBR). At these elevations the signal path length in the ionosphere is about three times that at high elevations. A typical narrow beam here can cover 200 to 300 kilometers of varying ionospheric structure, ten times the region covered by the same beam at high elevations. Also, the beam velocity relative to the structure will be much greater than when tracking the identical space vehicle observed at high elevation angles.

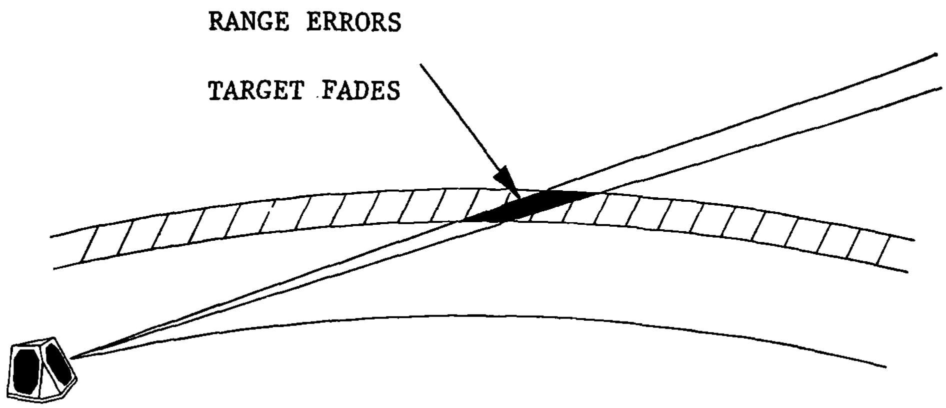


Figure 2. TRANS-IONOSPHERIC EFFECTS ON GROUND-BASED RADARS

Figure 3 shows essentially the reverse of the geometry of Figure 2, as it applies to the SBR application. At these low elevations, the conditions seen by the SBR may be very different from those above the target area. As seen in Figure 3, this geometry effect is beneficial, since the SBR is "looking under" a disturbed region. Obviously, the effect also can be detrimental, and care must be taken when near-real-time models specify the "location" of boundaries of disturbed regions.

The above examples of special geometries imply, at least, that great care must be taken in tailoring ionospheric effects products for such surveillance systems to assure that proper translation is made from model predictions of effects "near vertical" or effects as seen observing a geostationary platform. In fact, most existing empirical data is from near-vertical observations, and further work is being done to accumulate data on effects at low-elevation angles to validate model products for that geometry.

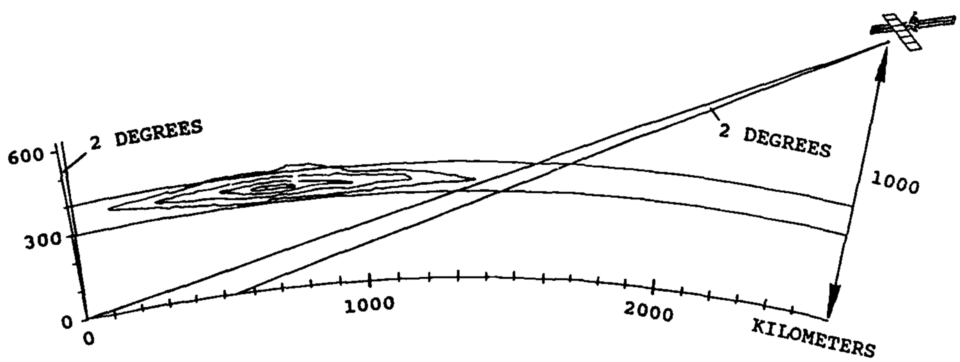


Figure 3. EXAMPLE OF SPACE-BASED RADAR/IONOSPHERE GEOMETRY

III. SPECIFICATION OF TRANS-IONOSPHERIC PROPAGATION EFFECTS

The United States Air Force's Air Weather Service has the responsibility for the U. S. Department of Defense in providing timely predictions of the Space Environment. The Space Forecasting Center, located in Colorado Springs, CO, will provide updated forecasts of various ionospheric parameters tailored to the needs of specific DoD customers who require correction for trans-ionospheric propagation effects. In order to provide these forecasts at the present time, the Air Force Global Weather Central Space Environmental Support Branch runs models of the ionosphere, updated with near-real-time measurements of the various solar, magnetic and direct ionospheric measurements which are required as model inputs. The models themselves are also undergoing continual, long term improvement by research work being done at the Air Force (Systems Command) Geophysics Laboratory and at other organizations.

One of the primary requirements of the models is for sufficient near-real-time data to update the average climatology which current models are capable of providing. In the case of trans-ionospheric propagation, a new sensing system, named TISS, for Trans-Ionospheric Sensing System, is currently under development for deployment by the early 1990s. TISS will provide the needed near-real-time data to allow the Space Forecasting Center to make greatly improved specifications of the state of trans-ionospheric propagation effects upon specific systems, and will allow realistic predictions to be made of future expected magnitude of the effects.

IV. CHARACTERISTICS OF THE TRANS-IONOSPHERIC SENSING SYSTEM

By monitoring signals from the NAVSTAR/Global Positioning System, (GPS), satellites the TISS will make multi-directional measurements of the specific trans-ionospheric propagation parameters which are of greatest importance to military systems. These parameters are: 1) Total Electron Content, (TEC), 2) the rate of change of TEC, 3) amplitude scintillation and 4) phase scintillation. Thus the TISS measures the major parameters of the ionosphere which can affect systems. Those effects which are proportional to TEC and its rate of change are directly obtained by the TEC measurement and its rate of change; the amplitude and phase scintillation parameters also are measured separately, and directly, by TISS.

A. CHARACTERISTICS OF THE GPS SATELLITES

The GPS satellites transmit spread spectrum signals, encoded so that they do not interfere with other users of the L-band spectrum, and are relatively immune to interference. In addition, without a knowledge of the modulation code, unauthorized users of the GPS cannot acquire the signals. All the GPS satellites utilize the same frequencies of 1228 MHz, called L2, and 1575 MHz, called L1, but each has its own unique modulation code which is orthogonal to all the other satellites. Hence, no mutual interference occurs between satellites.

There are to be 21 primary GPS satellites in 12 hour orbits at an inclination of 55 degrees. The orbital configuration of the GPS satellites was designed so that a minimum of four satellites would be in view from any ground station at all times. This orbital configuration allows trans-ionospheric propagation measurements to be made in a minimum of four directions continuously from each TISS ground station. A description of the characteristics of the GPS satellites is given by Denaro[9].

B. TEC MEASUREMENTS WITH TISS

Absolute TEC measurements are made by means of differential group delay of two, 10.230 MHz modulated, signals on the GPS satellites[10]. Any non-zero transmitted differential modulation phase, along with receiver and antenna differential phase errors, must be accounted for in the final measurement. Figure 4 illustrates data, taken with a prototype TISS monitor station, which shows the TEC along the slant path to four GPS satellites over a 4 hour period, from a station located at Sondrestrom Fjord, Greenland[11]. Values of satellite elevation and azimuth for the pass are also plotted in Figure 4. Note that the observed TEC varies significantly in the four different regions being observed by the prototype TISS receiving system.

It is the usual practice in making measurements of slant TEC to translate them to equivalent vertical values at a mean ionospheric height, generally taken to be between 350 and 450 km. The equivalent vertical TEC values are then assigned to the latitude and longitude of the point where the straight line from the satellite to the ground station intersects the mean ionospheric height. For observations at low elevation angles the earth angle from the ground station to the mean ionospheric penetration location can be as much as 15 degrees. That is, if one is making observations of the ionospheric TEC north or south from one ground station, equivalent vertical values can be obtained from latitudes as far away as plus and minus approximately 15 degrees from the latitude of the station. Equivalent distances in longitude also occur for satellites viewed east or west from an observing station.

The accuracy of the absolute value of TEC from TISS will be approximately one TEC unit, (1×10^{16} el/m^2,), but this level of accuracy cannot be obtained from the differential group delay only, due to the noise in the receiving system. Differential

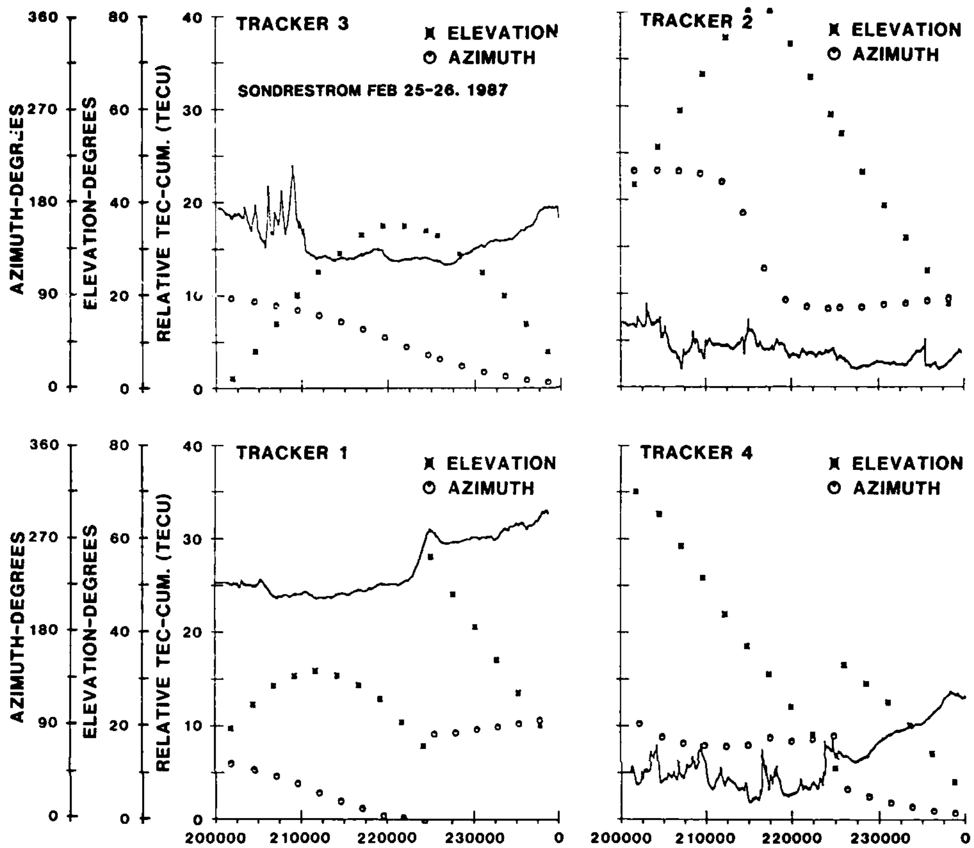


Figure 4. FOUR CHANNEL SLANT TEC VARIATION AND RAYPATH LOCATIONS

carrier phase measurements[10], however, can make <u>relative</u> TEC measurements to an accuracy of approximately $3X10^{14}$ el/m^2 in a bandwidth of 10 Hz, obviously much better than the <u>absolute</u> TEC measurements obtained from the differential group delay. The relative TEC measurement of the differential carrier phase, along with the poorer resolution, but excellent absolute TEC measurements, available from the differential group delay, are combined in a minimum root mean square error sense over some designated time interval. In this manner one has the best of both worlds. In the TISS, within approximately 15 minutes after initial lock on to a new satellite, the relative TEC from the differential carrier phase measurement can be fit to an absolute TEC scale to within approximately one TEC unit.

C. PHASE SCINTILLATION

The rate of change of TEC, measured from differential carrier phase on the two GPS transmitted carrier frequencies, also is used to determine phase scintillation. Phase scintillation is generally measured as a variation of phase about a detrended mean value over some standard detrend interval, generally one minute. TISS will determine and report the standard deviation of the measured differential phase, and the spectral strength and slope parameters for the power spectrum of the phase fluctuations, over each interval[12]. This will be done for all GPS satellites observed simultaneously, and the results will be referred to an equivalent single frequency of 1.0 GHz. The practice of referring the differential phase measurements observed at the L2 frequency minus that on the higher L1 frequency, to an equivalent single 1.0 GHz frequency is done so that any user can easily then refer the phase scintillation parameter to any required system frequency.

D. AMPLITUDE SCINTILLATION

TISS will determine amplitude scintillation directly from measurements of signal amplitude on the two GPS equivalent carrier frequencies. The standard measure of

amplitude scintillation is the r.m.s. deviation about the average signal level divided by the average signal level, expressed as a fraction. This value is called the scintillation index, S_4. TISS will derive and report S_4 for both of the GPS frequencies on each satellite observed over the same reporting intervals used for the phase scintillation parameters.

E. APPLICATION

The TEC and scintillation values obtained by TISS in several directions simultaneously are to be used to update models of the ionosphere, in the vicinity of each TISS station, from monthly median conditions, to the actual ionospheric parameters and gradient values measured from the TISS. In this manner each TISS site, monitoring dual frequency signals from GPS satellites in multiple directions, will ensure a large geographic coverage area of near-real-time updating of the monthly average model of the F region bulk plasma, discussed in section II above, as well as the monthly average model of scintillation, to be available at the Space Forecast Center.

V. PROPOSED DEPLOYMENT OF TISS

The TISS is a robust system designed to work in an unattended mode at remote locations and automatically send its measurements to the Space Forecast Center where the data is to directly update appropriate models for use in providing real-time products describing ionospheric effects. Specification and prediction of the appropriate trans-ionospheric propagation effects are then tailored to the needs of individual DoD customers. A map of proposed TISS sites is shown in Figure 5. The cross-hatched area around each station is the equivalent portion of the ionosphere which each TISS would be able to monitor by observing GPS satellites above 15° elevation. The "hole" in coverage which occurs for each station is due to the 55° inclination of the GPS satellites. Note that the stations closer to the geographic pole have this "coverage hole" almost overhead. Coverage of the North American Continent, especially in the important auroral and polar cap regions, is extensive, and is designed to provide the opportunity for near-real-time measurements of the ionospheric main trough, the scintillation boundary region, and the state of scintillation activity in the polar cap region. Latitudinal coverage is also provided from nearly the magnetic equator to the magnetic pole along the eastern part of the North American Continent.

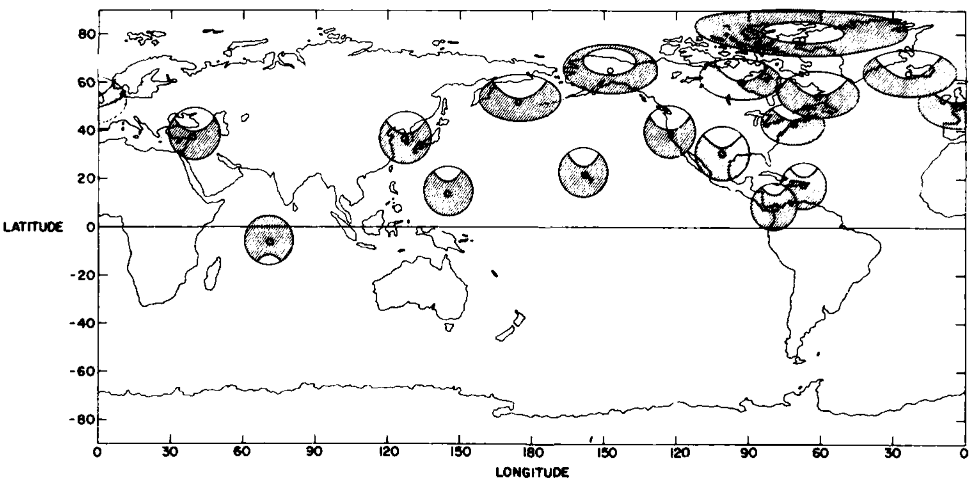


Figure 5. PROPOSED TISS STATION GEOGRAPHIC COVERAGE (350 KM ALT., 15° ELEVATION)

The TISS may be co-located, at several locations, with the AWS Digital Ionospheric Sounding System DISS, which is a system of modern high frequency digital ionospheric sounders designed to automatically make real time profiles of the electron density of the bottomside of the ionosphere. The combination of the bottomside shape, from the DISS, along with the TEC and the scintillation parameters from the TISS, makes a powerful combination of sensors providing complete real-time information on the state of the ionosphere.

VI. TISS AS A TACTICAL DECISION AID

The future combination of upgraded world-wide monthly median models of ionospheric effects residing at the Space Forecast Center, (SFC), capable of being updated with near-real-time inputs to the SFC from the TISS sites, will allow the SFC to provide

near-real-time products specifying ionospheric effects to DOD systems, world-wide. As required by AWS customers such products may be tailored specifically to provide the desired parameters in the appropriate geometry covering the required region for each system.

SFC's future capability to generate near-real-time reports will better enable future AWS operational system customers to adjust operating modes to mitigate ionospheric effects. Figure 6 charts a hypothetical flow of this process from SFC to the Command where the tactical decision would be made on the use of systems according to the reported situation. Table 6 tabulates a hypothetical process of how the ionospheric conditions could require changes in operational mode.

Figure 6. TISS AS A TACTICAL DECISION AID FOR A HYPOTHETICAL SYSTEM

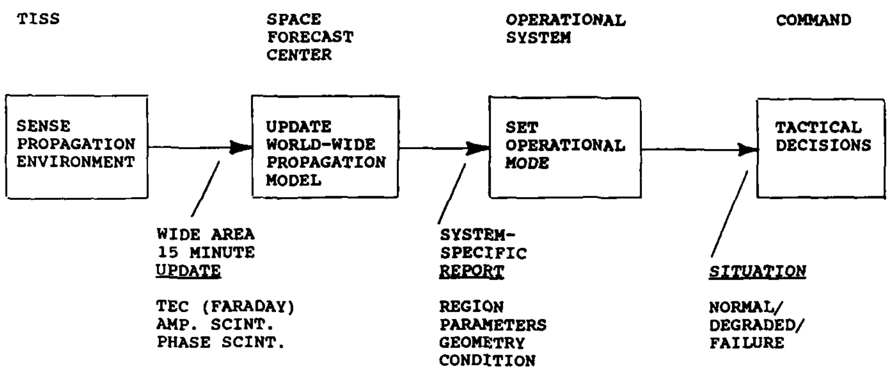


TABLE 6. **HYPOTHETICAL SYSTEM RESPONSE TO IONOSPHERIC CONDITIONS**

IONOSPHERIC CONDITION	OPERATION MODE	POTENTIAL DEGRADATION	REPORTED SITUATION
QUIET	NOMINAL OPERATION PRIMARY SENSOR	NOMINAL OPERATION	NORMAL
SIGNIFICANT DISTURBANCE	ALTERNATE SENSOR SPACE VECHILE (SBR) GROUND SITE (GBR) CHANNEL (COMM)	LOOK-AROUND POSSIBLE THRUPUT REDUCTION	DEGRADED
	ALTERNATE SIGNAL FORMAT /SIGNAL PROCESSING	MORE SYS RESOURCES RQD LESS THRUPUT	
	SENSE/MITIGATE ENVIRONMENT (if possible)	MAJOR SYS RESOURCES RQD THRUPUT IMPACTED SERIOUSLY DEGRADED MISSION	
SEVERE DISTURBANCE	ALTERNATE SYSTEM (if available)	DEGRADED MISSION MISSION FAILURE	FAILING

In addition to TISS's wide coverage area at each site, as ill· ' in Figure 5, which permits referencing of models with several measureme gradients of parameters at each site, the TISS will incorporate a directional capability on command. Since typically more than 4 GPS satellites will be visible from a given site, TISS will be capable of directing its observations in a given quadrant of the sky, or overhead, versus the normal mode, which seeks widely distributed coverage. The value of this capability is that TISS can concentrate observations in a region of more strategic significance within its coverage area, or when the wide coverage detects a significant event, such as the entry of auroral or equatorial disturbances into a given TISS coverage region, the observations may be concentrated to determine the extent and motion of the disturbed region.

To illustrate this potential capability, Figure 7 shows optical observations of airglow from ionospheric disturbances that occurred during the four hour observation period covered in Figure 4. The circled area and symbols in the figure show the location of ray paths to the four GPS satellites being observed. Two of these paths lie within one of the optical features, one at high elevation, and one at a low elevation. Similarly, the other two lie outside the feature with one each at high and low elevation angles. Figure 8 gives an expansion of plots in Figure 6, covering the times of the

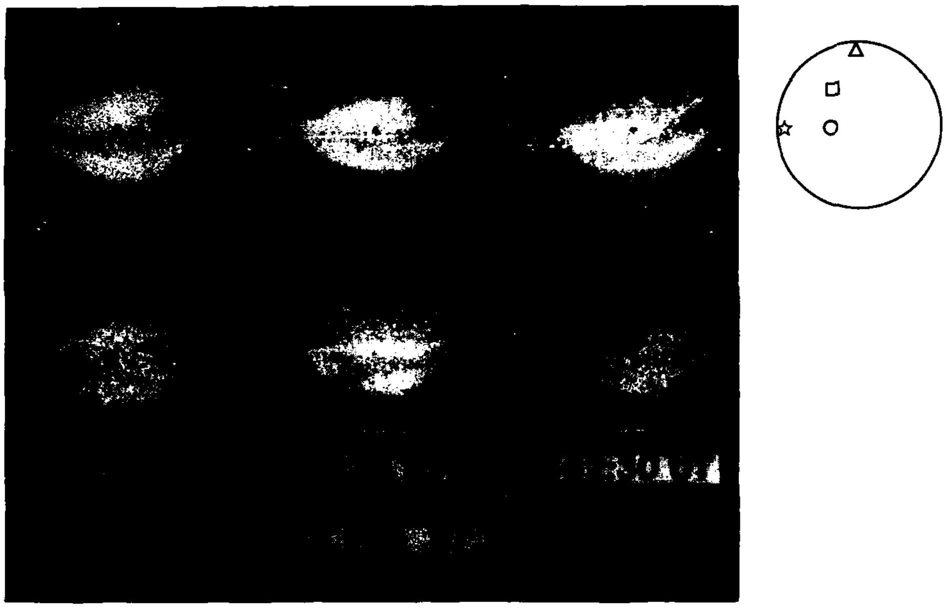


Figure 7. All-SKY IMAGES ILLUSTRATING IONOSPHERIC STRUCTURES OBSERVED AT SONDRESTROM, (SYMBOLS MARK RAYPATHS TO GPS SATELLITES)

optical observations. The two plots at the right side of the figure are from the paths inside the optical features. Higher levels of small scale disturbance are clearly visible in these plots compared with levels in the plots from the paths outside the observed optical feature. These effects were observed during very low solar activity. Subsequent data from near solar maximum conditions has shown much more dramatic differentiation between disturbed and quiet directions in the sky above a TISS site.

VII. CONCLUSIONS

The Trans-Ionospheric Sensing System, when fully deployed in the early 1990s, will be able to provide real-time updates to the DoD community for the effects of the ionosphere on trans-ionospherically propagated radio waves. The system is designed to be robust, fully automated, and to include first order artificial intelligence techniques to optimize data collection and interface with models.

VIII. REFERENCES

1. Basu, Santimay, E. MacKenzie and Sunanda Basu, "Ionospheric Constraints on VHF/UHF Communications Links During Solar Maximum and Minimum Periods", Radio Science, Vol. 23, No. 3, pp. 363-378, May-June 1988.

2. Klobuchar, J. A., "A First-Order, Worldwide, Ionospheric Time-Delay Algorithm", AFCRL-TR-75-0502, 25 Sept. 1975.

3. Klobuchar, J. A., G. J. Bishop, P. H. Doherty, "Total Electron Content and L-Band Amplitude and Phase Scintillation Measurements in the Polar Cap Ionosphere", AGARD-CPP-382, May 1985.

4. Whitney, H. E., J. Aarons, R. S. Allen, and D. R. Seemann, "Estimation of the Cumulative Amplitude Probability Distribution Function of Ionospheric Scintilla tions", Radio Science, Vol. 7, pp. 1095-1104, 1972.

5. Aarons, J., "Global Morphology of Ionospheric Scintillations", Proc. IEEE, Vol. 70, No. 4, pp. 360-378, April 1982.

6. Handbook of Geophysics and the Space Environment, A. S. Jursa, editor, National Technical Information Service, 5285 Port Royal Road, Springfield VA 22161, 1985

7. Tascione, T. F., H. W. Kroehl and B. A. Hausman, "ICED - A New Synoptic Scale Ionospheric Model", pp 299-309, in The Effect of the Ionosphere on Communication, Navigation, and Surveillance Systems, Library of Congress Cat. No. 87-619868, 1987.

8. Secan, J. A., E. J. Fremouw, R. E. Robins, "A Review of Recent Improvements to the WBMOD Ionospheric Scintillation Model", pp 607-616, in The Effect of the Ionosphere on Communication, Navigation, and Surveillance Systems, Librarv of Congress Cat. No. 87-619868, 1987.

9. Denaro, R. P. "Navstar: the All-Purpose Satellite", IEEE Spectrum, Vol. 18, pp 35-40. May 1981.

10. Klobuchar, J. A., "Ionospheric Effects on Earth-Space Propagation", AFGL-TR-84-0004., 27 Dec. 1983.

11. Bishop, G. J., D. J. Jacavanco, D. S. Coco, C. E. Coker, J. A. Klobuchar, E. J. Weber, P. H. Doherty, "An Advanced System for Measurement of Transionospheric Radio Propagation Effects Using GPS Techniques", AFGL-TR-87-0319, 13 Nov. 1987

12. Basu, Sunanda, Santimay Basu, R. C. Livingston, E. MacKenzie and H. E. Whitney, "Phase and Amplitude Scintillation Statistics at 244 MHz from Goose Bay Using a Geostationary Satellite", AFGL-TR-82-0222, 6 Aug. 1982.

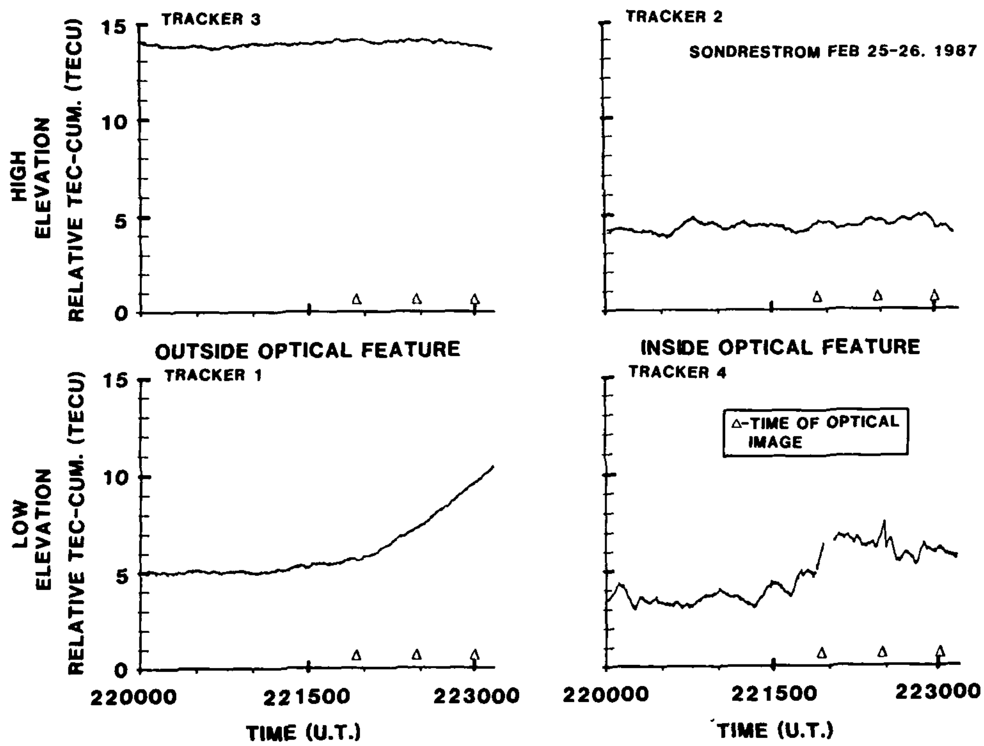


Figure 8. FOUR CHANNEL SLANT TEC VARIATION DURING TIMES OF OPTICAL IMAGES

A Portable Ionosonde in Support of Reliable Communications

D. Mark Haines
D. F. Kitrosser
B. W. Reinisch
University of Lowell
Center for Atmospheric Research
450 Aiken Street
Lowell, MA 01854

F. J. Gorman
U.S. Army Communications-Electronics Command
Fort Monmouth, NJ 07703

SUMMARY

The temporal and spatial variations in ionospheric structure have frustrated the efforts of communications and radar system operators who base their frequency management decisions on monthly mean predictions or computer based models. The University of Lowell Center for Atmospheric Research, in cooperation with the U.S. Army Communications-Electronics Command, has now produced a low power miniature version of its Digisonde sounders capable of making real-time on-site measurements to support communications or surveillance operations. The system compensates for a low power transmitter by employing intrapulse phase coding, digital pulse compression and Doppler integration. The data acquisition, control, signal processing, display, storage and automatic data analysis functions have all been condensed into a single multi-tasking, multiple processor computer system while the analog circuitry has been condensed and simplified by use of reduced transmitter power, wide bandwidth devices, and commerically available PC expansion boards.

Noteworthy new technology involved in this system includes:

a. An electronically switched active crossed dipole receiving antenna.

b. A TMS 320C25 digital signal processor.

c. Compact DC-DC converters allowing full operation on a single 24 VDC battery.

d. A high speed data acquisition board interfaced via an IBM-AT expansion bus.

f. A proprietary multi-tasking operating system.

g. Reprogrammable PROM based coherent oscillators.

h. Use of complementary codes which virtually eliminate the time domain pulse compression sidelobes typical of pulse compression systems.

i. Automatic mode identification and parameter scaling by an embedded expert system.

PREFACE

One of the chief applications for the real-time measurement data provided by current generation digital ionospheric sounders such as the Lowell Digisonde 256 (references 1 and 2) is to manage the operation of high frequency (HF) radio channels and networks. Since many HF radios are operated at remote locations (i.e. aircraft, boats, land vehicles of all sorts, and remote sites where telephone service is unreliable) the major obstacle to making practical use of the ionospheric sounder data and associated computed propagation information is the dissemination of this data to the operators' sites. Since HF is often used where no alternative communication link exists, or is held in reserve in case primary communication is lost, it is not practical to assume that centrally tabulated ionospheric data can be made available to the user. Furthermore, local measurements are superior to measurements at sites of opportunity in the user's general region of the globe since extreme variations in ionospheric properties are possible even over short distances, especially at high latitudes (reference 3), or a sunrise or sunset terminator. However, for most applications the size, weight, power consumption and cost of an ionospheric sounder have made local measurements impractical. Therefore the availability of a small, low cost sounder would be a major improvement in the usefulness of ionospheric sounder data. Shrinking the conventional 1 to 50 kW pulse sounders to a

portable battery operated 100 to 300 Watt system requires the application of substantial signal processing gain to compensate for the 10 to 23 dB reduction in transmitter power. Furthermore a compact portable package requires the use of highly integrated control, data acquisition, timing, data processing, display and storage hardware.

OBJECTIVE

The objective of the Portable sounder project (initiated and funded by the U.S. Army Communications-Electronics Command, Ft Monmouth, NJ) was to provide a vertical incidence (i.e. monostatic) ionospheric sounder which can automatically collect and analyze ionogram data at remote operating sites for the purpose of selecting optimum operating frequencies for obliquely propagated communication paths. Intermediate objectives assumed to be necessary to produce such a capability were the development of optimally efficient waveforms and of functionally dense signal generation, processing and ancillary circuitry. Since the need for an embedded general purpose computer was a given, as many functions as is feasible have been assigned to this computer rather than providing additional circuitry to perform these functions. The Portable sounder emulates most of the functions of the Digisonde 256 (DGS 256). These include the precise measurement of six observables at every sounding frequency:

(1) range (i.e. height)
(2) amplitude
(3) phase
(4) Doppler shift (e.g. finding the largest Doppler line)
(5) Doppler spectrum
(6) wave polarization.

A total characterization of the ionosphere is possible by measuring all observables at all detectable heights (maximum of 256 heights) and all propagating frequencies (typically about 70 to 200 frequencies). With multiple antennas and either an electronic antenna switch or multiple receivers the Portable could process and output angle of arrival information, a seventh observable which was not part of the project described here. To the maximum extent possible the operating modes and the data produced by the Portable system are identical to those of the DGS 256 so that it is compatible with automatic processing software already existing at ULCAR and at many other ionospheric research institutes worldwide.

SYSTEM FUNCTIONS

Some of the functions common to the Portable and the DGS 256 include:

1. Scanning Ionogram Programs (programs A, B or C) specify parameters for a stepped frequency ionogram mode. They have preassigned default parameters or can be modified by the user to select the start and stop frequencies, frequency step size, height resolution, and coherent integration time. Running this program produces a conventional ionogram such as in Figure 11.

2. Fixed Frequency Ionogram Programs are a continuous measurement at a single user defined frequency employing a programmable Doppler integration time and height resolution. This mode can be used to simultaneously detect changes in fading statistics, height, polarization, Doppler and phase of the ionospheric layers at a fixed frequency.

3. Automatic Schedules are stored records of desired start times for the different types of ionogram programs. Stop times are not programmed but occur naturally based on the selection of parameters in the ionogram program. Start times recur hourly but program schedules can be changed to a different set at programmable switchover times which can be specified by day number (1 to 366) and universal time, or can be specified to switchover at the same time every day (e.g. to run different ionogram parameters at night vs. daytime).

4. Discrimination of Ordinary and Extraordinary Polarization of the detected echoes is a major feature of these systems, made possible by rapidly switched left or right hand circularly polarized receive antennas (reference 4). The detection of O and X polarization is the key to successfully extracting ionogram characteristics with an automatic scaling program, which is done in real-time onsite.

5. Frequency Search - A technique which quickly scans several frequencies near the nominal next step in a scanning ionogram prior to transmission in order to select a frequency free of noise and interference.

6. *Full Doppler Spectrum* - Since the Portable performs the Doppler integration using a Fast Fourier Transform (FFT) all spectral lines (i.e. if 256 pulses are Doppler processed, 256 complex spectral lines are produced) are available at each height measured.

7. Automatic Scaling by the ARTIST (Automatic Real-Time Ionogram Scaler with True height analysis) program is performed on-site by the general purpose computer embeded in the Portable system. This program identifies traces formed by echoes from the E, sporadic E, F1 and F2 layers. The traces are then "inverted" to produce the inferred electron density profile (reference 5), which also can be printed out on the processed ionogram as true height of the plasma frequency.

8. Remote control and remote data access is provided via telephone lines and modems to a remote laptop portable computer. The remote can be used to reprogram schedules, to change program parameters and operating modes, or to query the system for and print out the most recent ionogram, the last 24 hours of foF2's or many other data sets.

THEORY OF OPERATION

During the 1960's and 1970's several variations in sounding techniques started moving significantly beyond the basic pulse techniques developed in the 1930's. First was the integration of several pulses transmitted at the same frequency. This increases the signal amplitude coherently while noise and other uncorrelated signals integrate incoherently giving an increase in signal to noise power of N^2/N when N pulses are processed. For coherent processing the integration time is limited to the interval over which the signal undergoes a phase shift of 90 degrees, unless a full Doppler integration (Fourier transform) is performed on the time domain signals, which allows spectral domain coherent integration of several hundred pulses (reference 4). Another technique FM/CW or chirp sounding (reference 6) was the most radical departure from pulse sounding since it involves a 100% duty cycle transmission and a great deal of signal processing gain. Since the transmitter is always on, the receiver and transmitter must be separated by a sufficient distance to avoid overloading the receiver amplifiers and mixers, typically several kilometers, but more often these systems are used for oblique propagation over 100's to 1000's of kilometers. The linear frequency sweep (e.g. 100 kHz/sec) converts propagation delays to frequency offsets, therefore the height of the ionospheric reflections is directly proportional to the frequency offset between the received signal and the current transmitted signal. The frequency difference is on the order of 70 to 500 Hz (corresponding to 0.7 to 5 milliseconds) and can be detected by spectrum analyzing the received baseband. A third general technique is to stretch out the pulse thus increasing the duty cycle so it contains more energy, but modulate it with an internal phase code to retain the height resolution of a shorter pulse (reference 7). The critical factor here is the correlation properties of the internal phase code which could be a Barker (reference 8), Huffman (reference 9), Convoluted Codes (reference 11), any of several "maximal length" shift register codes (reference 12) or Golay's Complementary Series code pairs (references 10 to 13). The internal phase code alternative has just recently become economically feasible with the availability of very fast microprocessor and signal processor IC's. Another new development in the 1970's was the coherent multiple receiver array (reference 4) which allows angle of arrival/incidence angle to be deduced from phase differences between antennas by standard interferometer techniques. Interferometry is invalid, however, if there are multiple sources contributing to the received signal. This problem can be overcome as in the DGS 256 by first isolating or discriminating the multiple sources before applying the interferometry relationships, but this technique is beyond the scope of the current paper.

Except for the FM/CW sounder which operates well on 10 to 100 watts (peak and average transmitter power), the above techniques and cited references typically employ a 5 kWatt to 50 kWatt peak power pulse transmitter. This power is needed to get sufficient signal strength to overcome an atmospheric noise environment which is typically 20 to 50 dB above thermal noise. Furthermore, since ionogram measurements require scanning of the entire propagating band of frequencies in the 0.5 to 30.0 MHz RF band, the sounder receiver will encounter broadcast stations, ground to air communications channels, HF radars, ship-to-shore channels and several very active radio amateur bands which can add as much as 60 dB more unwanted signal. Therefore, the sounder signal must be strong enough to be detectable in the presence of these large interfering signals. A coherent pulse sounder must have a broad receiver bandwidth to maintain the capability to accurately measure heights. This broad receiver bandwidth also complicates the move to lower transmitted power since the received external noise/interference, which is the largest noise source in the system, increases proportionally with increased bandwidth.

The Portable transmits only 300 Watts of pulsed RF power but compensates for this by long Doppler integration (up to 512 pulses) and by digital pulse compression, achieving a total of approximately 30 dB of signal processing gain. The signal processing incorporates two of the advances of the 1960's and 1970's by using both pulse compression and Doppler integration. The pulse compression allows using a longer transmitted pulse, which increases the signal energy, while maintaining the height resolution of a shorter pulse. In order to have a monostatic system the transmitted pulse is turned off by the time the first E-region echoes arrive at the receiver which is about 700 µsec after the beginning of the pulse (see Figure 1). Also the pulse repetition frequency is limited by the longest delay of interest which is at least 4.5 msec, corresponding to a 600 km propagation delay plus 500 µsec to receive the coded pulse. We selected a 533 µsec pulse made up of eight 66.67 µsec phase code chips which allows detection of ionospheric echoes starting at 80 km altitude. We chose a highest pulse repetition frequency of 200 pulses/sec which allows reception of the entire pulse from 670 km altitude before the next pulse is transmitted. This timing captures all but the higher multihop F region echoes which are of little interest anyway. When longer receive intervals are desired, the system can be operated at a 100 or 50 Hz pulse repetition frequency.

The received signal is usually a superposition of several phase coded echoes reflected at various ranges as depicted in the computer simulation of an eight chip Complementary Series waveform shown in Figure 2. The overlapping signals can be resolved by pulse compression processing (Figure 3) but energy reflected from any given height will leak or spill into other heights to some degree as a result of channel induced Doppler, imperfections in the phase code and/or imperfections in the phase response of the transceiver. Several codes were simulated and analyzed for leakage from one height to another and for tolerance to signal distortion caused by filtering. The pulse compression algorithm is a cross-correlation of the received signal with a unit amplitude replica of the code known to have been sent. Therefore it is the leakage properties of the autocorrelation functions which are of interest. The results of a simulation program run on a VAX computer for several different codes are shown in the following figures:

a. Complementary series (Figure 3)

b. Periodic M-codes (Figure 4).

c. Non-periodic M-codes (Figure 5).

d. Barker code (Figure 6).

e. Kasami sequence (Figure 7).

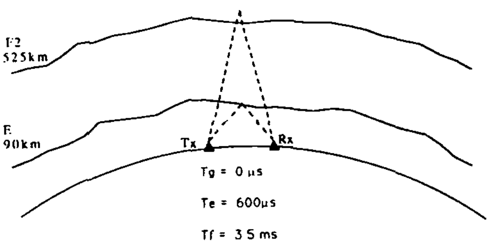


Figure 1. Time Window for Monostatic Transmission and Reception

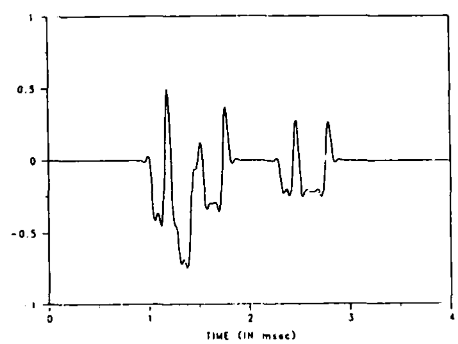


Figure 2. Simulated Reception of Band Limited Complementary Coded Pulses after Multipath Propagation

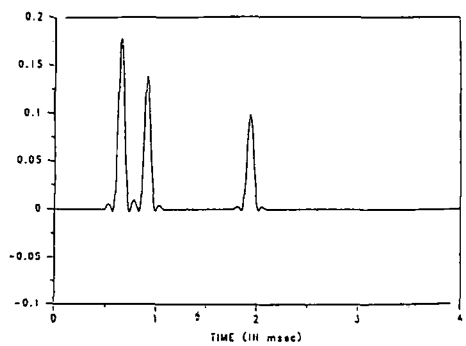


Figure 3. Simulated Complementary Coded Pulses after Pulse Compression Processing

Since only the Complementary series provides a mathematically perfect pulse compression (i.e. no energy is leaked into other heights) we selected this phase code scheme for the Portable. It does so by summing the pulse compressions (cross correlation functions) of two different codes transmit-

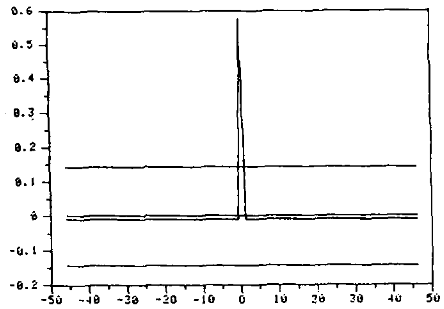


Figure 4. Autocorrelation of Periodic Maximal Length Coded PSK/CW Waveform, no Pulse Shaping Applied 63 Chips

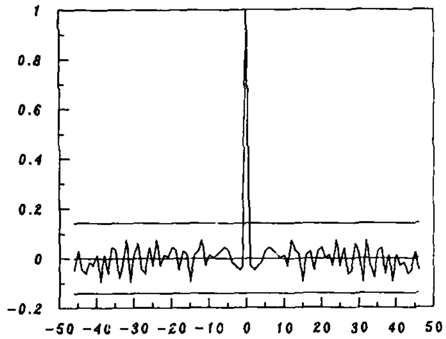


Figure 5. Non-Periodic Maximal Length Sequence Autocorrelation Function

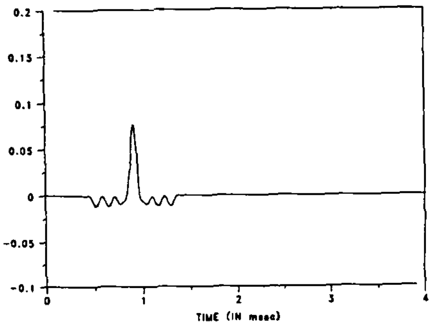


Figure 6. Band Limited Barker Coded Pulse after Pulse Compression

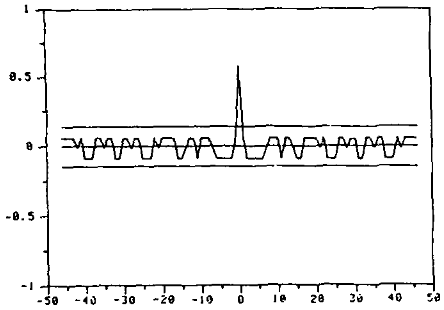


Figure 7. Autocorrelation of 63 Chip Kasami Code

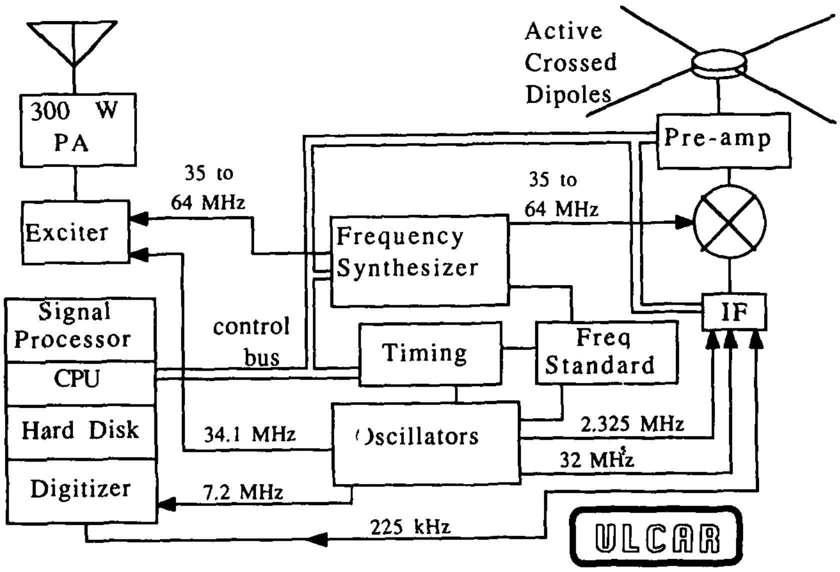


Figure 8. Portable Sounder System Diagram

tions) of two different codes transmitted successively which have exactly the inverse leakage of each other. Therefore the leakage at spurious heights cancel out; however this technique is sensitive to high Doppler shifts because *it requires phase stationarity* over a period of 5 msec (the interval from one pulse to the next). This could be a concern but the low leakage properties of the code are maintained "fairly well" up to ±25 Hz Doppler shifts. "Fairly well" means that the largest code leakage (spurious amplitude leaking into a neighboring range bin) is at one point 5 heights away from the actual height where the leakage is 19 dB below the peak. All other heights are contaminated by less than -28 dB at 25 Hz Doppler.

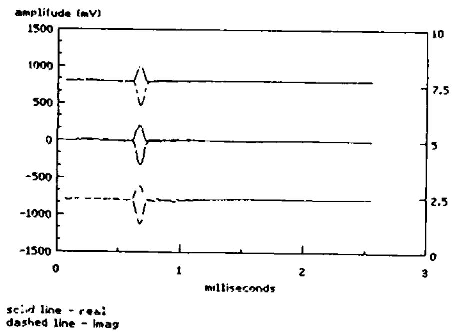


Figure 9. Received Pulses In Loopback Mode

Figure 8 is a block diagram of the Portable system. Figure 9 shows the result of the pulse compression performed with this system in a loopback self test configuration (i.e. the propagation channel was 2 m of coax cable and an attenuator). The code leakage due to signal distortion in the system was a maximum of 27 dB below the compressed pulse but some of this spurious response is leakage of the transmitter's local oscillator which is normally off during the reception period (note the pedestal on which the compressed pulse sits). In actual operation we don't see code compression sidelobes until there is a SNR of 36 dB.

SYSTEM CONFIGURATION

Except for the Zenith Z-184 laptop portable computer (for human interface), the batteries, the external frequency synthesizer (either a PRC-104 military HF radio or a parallel BCD controlled commercial frequency synthesizer), and the transmit and receive antennas, the entire configuration of the portable is shown in Figure 10. Some specifics about system components are:

1) The receive antennas are electronically switched from right-hand to left-hand circular polarization between pulses by control signals which are multiplexed with the DC power and the RF signal on a single coaxial cable.

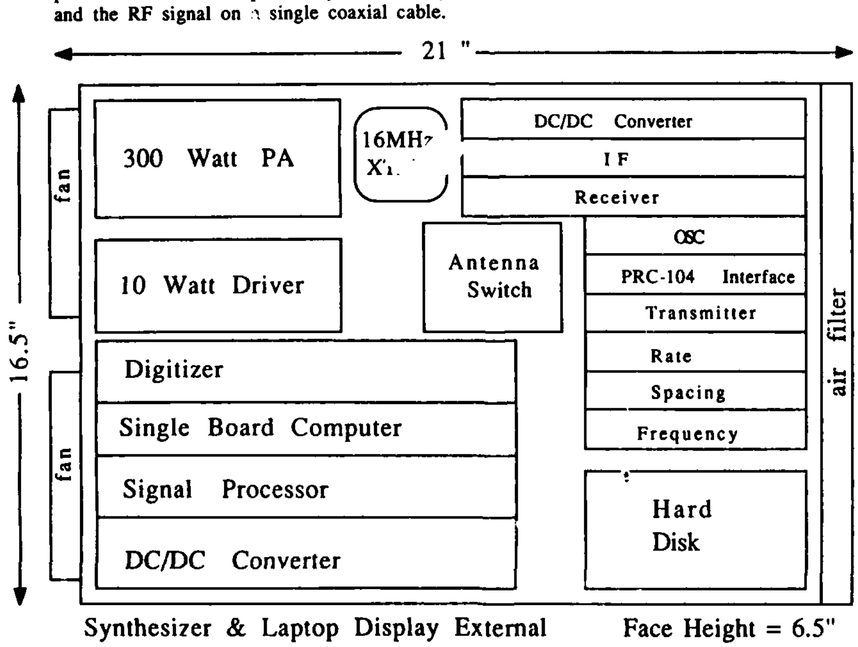


Figure 10. Portable sounder chassis layout

2) The double conversion receivers employ electronically tuned front end filters of 5% to 10% bandwidth to improve sensitivity.

3) The A/D converter is a high speed complex sampling 12-bit circuit built on a PC board installed in the computer backplane.

4) The signal processor is a commercially available board based on the Texas Instruments TMS320C25 (a 10 million instructions per second digital signal processor) which performs the Doppler integration at each range using an FFT algorithm.

5) The CPU and all input/output control is provided by an IBM AT compatible single board computer which is located in the system chassis.

6) DC/DC converters allow the entire system to be powered by two 12 volt batteries.

The ionograms resulting from the portable can be stored on magnetic disk or printed on site. The automatically scaled ionogram in Figure 11 was made with the Portable sounder, which was developed for the US Army Communications - Electronics Command. The amplitude, wave polarization, and binary Doppler shift (i.e. positive or negative) can be visually read from the printed ionogram as well as the standard height versus frequency electron density profile. The automatically scaled ionogram parameters are printed within one minute after the end of the ionogram transmission. For the Army's application it was desirable to compute the MUF value for various range communications channels. These values for 50 through 600 km are displayed along with the layer heights and critical frequencies in the text portion of the ionogram printout.

TEST RESULTS

A two day continuous test was undertaken during which we ran a DGS 256 side-by-side with the Portable. The DGS 256 transmitted 10 kW peak power on a 160 ft (50 m) vertical rhombic antenna and received on 7 in-phase (i.e. forming a vertical beam) magnetic loop receiving antennas. This configuration ensured a high quality standard against which to measure the performance of the Portable. In comparison, the Portable transmitted 150 W peak RF power on a 40 ft (12 m) high vertical log-periodic antenna and received on a single active crossed dipole antenna (2 m dipole length). The foF2 values automatically scaled on the Portable system during that period were compared with manually verified values from the DGS 256. Although the Portable's ionograms were noticeably weaker than the DGS 256's, correct scaling (within 0.5 MHz) was achieved in 81% of the ionograms. Furthermore, the majority of these errors were due in part to the automatic scaling program since

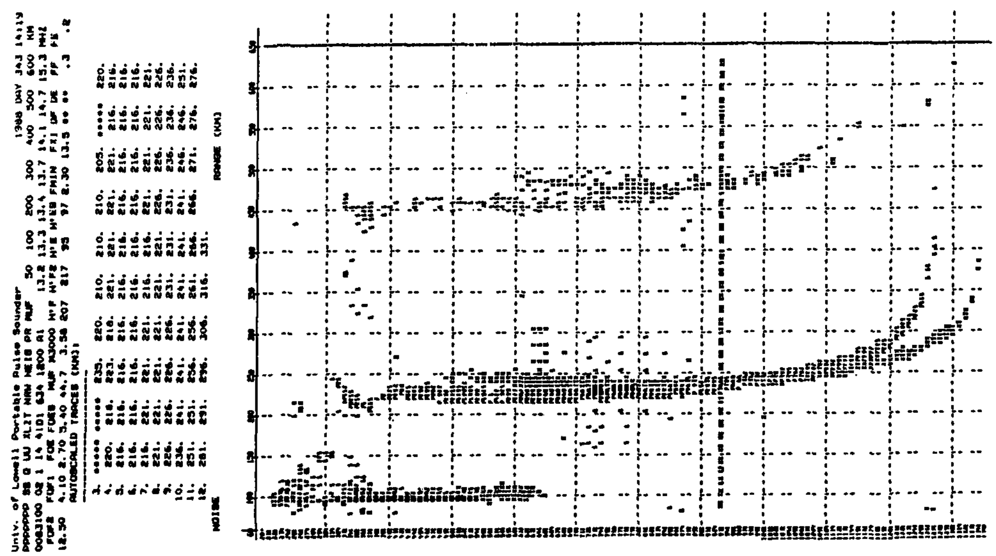


Figure 11. Scaled Portable HF Pulse Sounder Ionogram

Dec. 9, 1988, Westford, MA, 14:19 Z

manual scaling of the foF2 from the portable system's data was possible in all but 2% of the cases. Some major improvements in the scaling program are currently underway which directly address the problem of scaling weak or broken traces correctly.

We also compared performance of our recently developed transportable vertical log-periodic antenna with that of a horizontal folded dipole and with the 50 m vertical rhombic antenna. The rhombic provided 6-8 dB higher signal to noise ratios than the vertical LPA while the dipole was equivalent to the vertical LPA at its resonant frequencies but had deep nulls near its antiresonant frequencies.

REFERENCES

1. Bibl, K., B. W. Reinisch and D. F. Kitrosser (1981), "General Description of the Compact Digital Ionospheric Sounder Digisonde 256", Univ. of Lowell Center for Atmospheric Research
2. Reinisch, B. W. (1986), "New Techniques in Ground-Based Ionospheric Sounding and Studies," Radio Science, Vol 21 (3), May-June.
3. Buchau, J., B. W. Reinisch, E. J. Weber and J. G. Moore (1983), "Structure and Dynamics of the Winter Polar-Cap F-Region," Radio Science, 18(6), 995.
4. Bibl, K. and B. W. Reinisch (1978), "The Universal Digital Ionosonde", Radio Science, Vol. 13, No. 3, May-June.
5. Reinisch B. W. and X. Huang (1983), "Automatic calculation of electron density profiles from digital ionograms," Radio Science, Vol 18, No. 3, May-June.
6. Barry, G. H. (1971), "A Low Power Vertical Incidence Ionosonde", IEEE Transactions on Geoscience Electronics, GE-9, No. 2, 86-89.
7. Grubb, R. N. (1979), "The NOAA SEL HF Radar System (Ionospheric Sounder), NOAA Technical Memorandum ERL SEL-5
8. Barker, R. H. (1953), "Group Synchronizing of Binary Digital Systems", Communication Theory, London, pp. 273-287.
9. Huffman, D. A. (1962), "The Generation of Impulse-Equivalent Pulse Trains, IRE Transactions on Information Theory," IT-8, September.
10. Schmidt, G., R. Ruster, and P. Czechowsky (1979), "Complementary Code and Digital Filtering for Detection of Weak VHF Radar Signals from the Mesosphere," IEEE Transactions on Geoscience Electronics, May.
11. Coll, D. C. (1961), "Convoluted Codes," Proceedings of the IRE, Vol 49, No. 7.
12. Sarwate, D. V. and M. B. Pursley (1980), "Crosscorrelation Properties of Pseudorandom and Related Sequences," Proceedings of the IEEE, Vol. 68, No. 5, May.
13. Golay, M. J. (1961), "Complementary Codes," IRE Transactions on Information Theory, April.

CORRECTION OF IONOSPHERIC EFFECTS FOR THE PRECISE ORBIT DETERMINATION OF SATELLITES

F. Foucher, R. Fleury, P. Lassudrie-Duchesne
Centre National d'Etudes des Télécommunications
LAB MER/GER, 22301 Lannion, France

SUMMARY

In a near future, space missions in altimetry, geodesy or precise positioning will often require satellite orbits to be determined with subdecimetric accuracies. In orbit determination systems extensive use is made of Doppler measurements performed on radio signals propagating between the satellite and ground stations. Such systems, however, are highly sensitive to propagation errors brought about by the atmosphere. In this paper the effects of the ionosphere on Doppler measurements is assessed and a model of the ionospheric error is described. A correction scheme is examined for orbit determination systems based on Doppler data. Particular emphasis is placed on the post-processing of the Doppler data from the DORIS system, a precise orbit determination system designed for remote sensing satellites of the new generation. Simulation results show that the major part of the ionospheric error can be removed by model correction.

1. INTRODUCTION

Satellite orbits are usually derived from Doppler frequency shift measurements performed on radio signals propagating between ground stations and the satellites. Doppler measurements are however affected by these parts of the atmosphere where the refractive index is different from unity, i.e. the troposphere and the ionosphere. This paper deals with the assessment and correction of ionospheric errors on the Doppler data used in orbit determination systems. The ionospheric error can be conveniently expressed as a power series expansion in the inverse of the frequency. By using two-frequency systems, the first order ionospheric error term is usually eliminated. However, the relative weight of the higher order terms is then increased. The latter terms (mainly the 2nd and 3rd order terms) can be compensated for by model correction *[Clynch et al., 1979]*, *[Lassudrie-Duchesne et al., 1986]*. The inputs to the correction model are values of the Total Electron Content (TEC) of the ionosphere defined as the number of free electrons in a column of ionosphere parallel to the ray path and of unit cross-section. Also examined is an algorithm, currently under test, aimed at deriving TEC values from two-frequency Doppler systems. Particular emphasis is placed on the post-processing of the DORIS Doppler data. The DORIS system will be a precise orbit determination system consisting of about 50 transmitting ground stations together with a space borne platform that will collect dual frequency (400/2000 MHz) Doppler data and transmit them back to earth. The DORIS system is to be operated on remote sensing satellites of the new generation, among them are the imaging satellite SPOT-2 and the French-US ocean altimetric satellite TOPEX/POSEIDON.

2. DOPPLER SHIFT FOR TRANSIONOSPHERIC PROPAGATION

A general expression for the Doppler shift of a CW wave propagating on a transionospheric path is :

$$\Delta f = -\frac{f}{c}\frac{dP}{dt} \quad ; \quad \text{with} \quad P = \int_S n\, ds \qquad (1)$$

where c is the speed of light in vacuum, P the phase path length, n the real part of the refractive index of the ionosphere and S the electromagnetic ray path. As a consequence of the satellite motion, S is a function of the time t.

Denoting by $\mathcal{L}$ the straight-lined path between the satellite and the ground station, the phase path length can be expressed as :

$$P = \int_{\mathcal{L}} d\ell + \int_{\mathcal{L}} (n-1)d\ell + \left(\int_S n ds - \int_{\mathcal{L}} n d\ell\right)$$

or

$$P = L_0 + \int_{\mathcal{L}} (n-1)d\ell + R \qquad (2)$$

L_0 is the true range (assuming that corrections have been made for all other biases, e.g. clock biases and tropospheric delays). The second term in eq.(2) arises as a consequence of the refractive index being different from unity. The term R is due to the curvature of the ray path. The refractive index of the ionosphere is given by the Appleton-Hartree formula *[Budden, 1966]*. For frequencies above 100 MHz, it can be expanded into a power series in $1/f$:

$$n = 1 - \frac{X}{2} + \eta\frac{X|Y_L|}{2} - \frac{X^2}{8} - \frac{X\,Y_L^2}{2} + O\left(\frac{1}{f^5}\right) \qquad (3)$$

with $X = kN_e/f^2$ and $Y_L = (f_g/f)\cos\theta$; N_e is the electron density of the ionosphere ; f_g, the gyrofrequency of the electrons in the geomagnetic field B_0 ; θ, the angle between the geomagnetic field and the wave vector ; $\eta = \pm 1$ according to the propagation mode. Inserting eq.(3) into (2) yields :

$$P = L_0 + \frac{\ell_1}{f^2} + \eta\frac{\ell_2}{f^3} + \frac{\ell_3}{f^4} + R + O\left(\frac{1}{f^5}\right) \qquad (4)$$

with

$$\ell_1 = a\int_{\mathcal{L}} N_e\, d\ell = a\, I \quad ; \quad I = \int_{\mathcal{L}} N_e\, d\ell$$

$$\ell_2 = b\int_{\mathcal{L}} N_e B_0 |\cos\theta| d\ell \quad ; \quad \ell_3 = c_1\int_{\mathcal{L}} N_e^2 d\ell + c_2\int_{\mathcal{L}} N_e B_0^2 \cos^2\theta\, d\ell$$

$$R = \int_S n d\ell - \int_{\mathcal{L}} n d\ell$$

where a, b, c_1 and c_2 are constants. I is the ionospheric Total Electron Content (TEC) defined as the number of free electrons in a column of ionosphere parallel to $\mathcal{L}$ and of unit cross section. The terms ℓ_1 , ℓ_2 , ℓ_3 and R appear as perturbation terms to L_0 induced by the ionosphere. Typical values of these terms are given in *fig. 1*.

The Doppler shift can be deduced from eqs.(1) and (4) :

$$\Delta f = -\frac{f}{c}\frac{dL_0}{dt} - \frac{1}{cf}\frac{d\ell_1}{dt} - \frac{\eta}{cf^2}\frac{d\ell_2}{dt} - \frac{1}{cf^3}\frac{d\ell_3}{dt} - \frac{f}{c}\frac{\partial R}{\partial t} + O\left(\frac{1}{f^4}\right) \qquad (5)$$

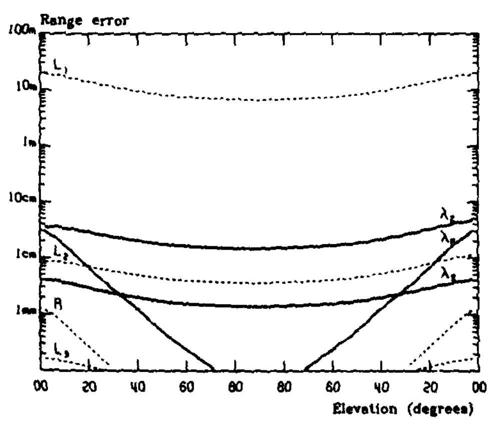


Fig. 1. Magnitudes of the phase path perturbation terms of ionospheric origins $L_n = \ell_n/f^{n+1}$ for f = 2 GHz (TEC = 66×10^{16} m^{-2}). L_1, L_3 and R are negative. The sign of L_2 depends on the polarization of the received waves. λ_2, λ_3 and λ_R are given for the DORIS dual-frequency system operated at 400/2000 MHz.

The first term in this expression arises solely from the satellite motion : it is the Doppler shift that would be observed if the ionosphere were removed. The terms next to it are perturbation terms of orders 1, 2 and 3 respectively. It can be shown by simulation that the R -dependant term is of order greater than 3. The first order term is seen to be proportional to the time derivative of the TEC. As will be shown in the following, the latter parameter can be measured by using two-frequency systems.

In two-frequency orbit determination systems, Doppler shifts Δf_u and Δf_ℓ are measured at the upper and lower frequencies f_u and f_ℓ respectively. Expressions for Δf_u and Δf_ℓ can be obtained by truncating eq.(5) to the first order :

$$\left.\begin{aligned} \Delta f_u &= -\frac{f_u}{c}\frac{d\hat{L}_0}{dt} - \frac{1}{cf_u}\frac{d\hat{\ell}_1}{dt} \\ \Delta f_\ell &= -\frac{f_\ell}{c}\frac{d\hat{L}_0}{dt} - \frac{1}{cf_\ell}\frac{d\hat{\ell}_1}{dt} \end{aligned}\right\} \tag{6}$$

where $\hat{L}_0$ et $\hat{\ell}_1$ are first order estimates of L_0 and ℓ_1.

The differential Doppler Δf_d is defined by :

$$\Delta f_d = \Delta f_u - \frac{f_u}{f_\ell}\Delta f_\ell \tag{7}$$

Putting eq.(6) into eq.(7) yields :

$$\Delta f_d = \frac{C_d}{a}\frac{d\hat{\ell}_1}{dt} = C_d\frac{d\hat{I}}{dt} \tag{8}$$

where $\hat{I}$ *is a first order estimate of the TEC and* $C_d = \frac{a}{cf_u}\left(\frac{f_u^2}{f_\ell^2} - 1\right)$.

Eq.(8) shows that the time derivative of the TEC can be estimated from two-frequency Doppler measurements.

3. TEC EVALUATION FROM DOPPLER MEASUREMENTS

3.1. Basis of the method

As will be exemplified below, the TEC is a key parameter for transionospheric propagation models. The TEC along any ground station to satellite path can be calculated from eq.(8) only to within an unknown integration constant. This constant cannot be determined unless more data are used *[Leitinger et al., 1975]* or some assumption is made on the space variations of the TEC *[Lassudrie-Duchesne, 1986]*. In this section a method is described to evaluate the TEC in the vicinity of the satellite track when a sufficiently dense network of Doppler ground stations can be used. The method makes use of reasonable assumptions on the TEC space variations while requiring only limited overlap regions between contiguous stations.

A portion of the satellite orbit is depicted schematically in *fig. 2*. For the sake of clarity, we shall assume the orbit to be circular with a non zero inclination angle. The latitude of the satellite θ_s can thus be used as a parameter for the satellite position. Doppler ground stations G_i $(i = 1,\dots,p)$ are assumed to track the satellite during time intervals $[t_{1i}, t_{2i}]$. For each station, eq.(8) takes the form :

$$\frac{d}{dt}[I_{si}] = \frac{1}{C_d}\Delta f_{di} \tag{9}$$

where I_{si} is the slant TEC (estimated to the first order in 1/f) along the path S-G_i and Δf_{di}, the Doppler shift for this path.

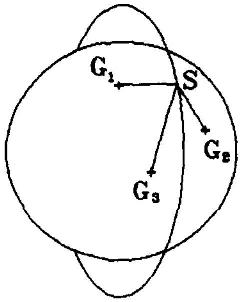


Fig. 2. Satellite orbit with ground stations (schematically). Data from stations yielding overlapping ionospheric traces are processed together in order to derive the TEC.

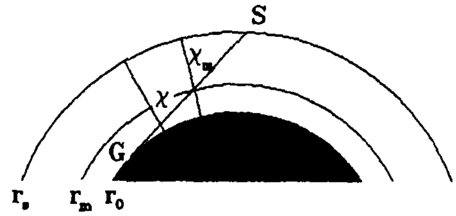


Fig. 3. Conversion of slant TEC to vertical TEC by means of the obliquity factor $\sec\chi_m$. The slant TEC along *G-S* is converted into a vertical TEC at the subionospheric point of the path of altitude r_m. The zenith angle at this point is χ_m.

The slant TEC can be converted into vertical TEC by means of an obliquity factor *[Leitinger et al., 1975]* defined as :

$$K = \frac{I_s}{I_v} = \int_{r_0}^{r_s} N_e \sec\chi \, dr \Big/ \int_{r_0}^{r_s} N_e \, dr \qquad (10)$$

where r_s is the geocentric altitude of the satellite, r_0 the earth radius and χ is the zenith angle of a current point along the path *(fig. 3)*. K depends in general on the variations of N_e along this path. However, simulations show that a mean value can be used for the $\sec\chi$ factor in eq.(10) which corresponds to the zenith angle χ_m of a particular point on the path refered to as the subionospheric point. The altitude r_m of the subionospheric point *is found to be roughly constant* : $r_m = r_0 + 400$ km. The vertical TEC defined by eq.(10) is then calculated at the position of the subionospheric point. Thus, for the *i*-th path a subionospheric point of coordinates (r_m, θ_{mi}) is defined and the slant TEC along the path can be converted to a vertical TEC at the latitude $\theta = \theta_{mi}$ by :

$$I_{si} = K_i . I_{vi} \quad ; \quad \text{with} \quad K_i = \sec\chi_{mi} \qquad (11)$$

Solving eq.(9) together with eq.(11) yields :

$$I_{vi}(\theta) = a_i(\theta) + b_i(\theta).c_i \qquad (12)$$

where θ denotes the latitude at the *subionospheric* point, c_i an unknown integration constant, $a_i(\theta)$ and $b_i(\theta)$ coefficients defined by :

$$a_i(\theta) = \frac{1}{C_d \; K_i(\theta)} \int_{t(\theta_{0i})}^{t(\theta)} \Delta f_{di} \; d\tau$$

$$b_i(\theta) = \frac{K_i(\theta_{0i})}{K_i(\theta)} \qquad (13)$$

with θ_{0i} an arbitrary initial value for the latitude. The solutions given by eq.(12) hold only for $\theta \in \Theta_i$, with Θ_i the interval of subionospheric point latitudes corresponding to the observations from the *i*-th station. Let us now make the assumption that the vertical TEC is approximately *invariant in longitude over distances typical* of the separation between the subionospheric points and the satellite track, i.e. a few hundred kilometers on the average. With this assumption, we can define a unique function I_v that represents the TEC in the vicinity of the satellite track. The function I_v must be such that :

$$I_v(\theta) = I_{vi}(\theta) \quad \text{for } \theta \in \Theta_i \; , \; (i = 1,\ldots,p) \qquad (14)$$

In practise, we shall restrict θ to a latitude interval Θ made up of overlapping intervals of observation Θ_i *(fig. 4)*.

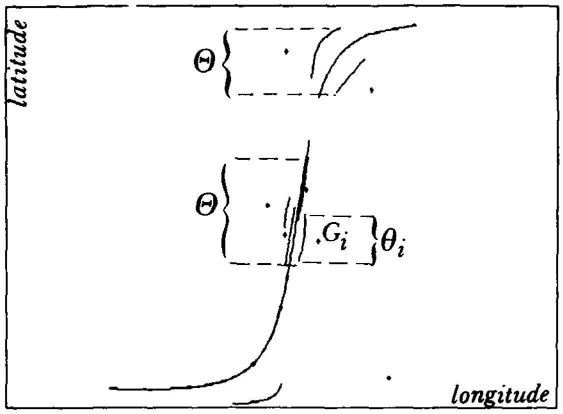


Fig. 4. Definitions of the intervals Θ_i and Θ. The subionospheric points corresponding to the i-th station belong to a latitude interval Θ_i. A group of overlapping Θ_i form an interval Θ. The TEC is derived on the whole interval Θ by processing together all the data from the same group.

Since the coefficients c_i are unknown, I_v will in general be undetermined unless some assumption can be made upon it. We shall therefore assume that I_v is a derivable function of θ with a continuous derivative I_v'. Since a number of measurement and modelling errors have been neglected so far, I_v cannot be determined in an exact manner. It is shown in the *Appendix* that a function $\tilde{I}_v$ can be found that approximates I_v in a space of finite dimension.

3.2. Simulation results

In order to test the TEC evaluation algorithm, simulations of Doppler measurements were carried out by using the expected trajectories of SPOT-2 and the planned DORIS ground station network. TEC data were simulated by integrating density profiles from the Bent model *[Llewellyn and Bent, 1973]*. Doppler data were computed for successive satellite positions with a repetition rate of 30 seconds (the actual repetition rate of the DORIS measurements will be about 10 seconds). The TEC evaluation algorithm was then applied to these data and values for the calculated TEC $\tilde{I}_v$ were derived. These values were then compared to the actual vertical TEC I_v that was found at the subionospheric points and an evaluation error was derived. This process was repeated for each group of overlapping intervals of subionospheric points. An overall error was finally computed for the whole satellite pass.

Results are shown in *fig. 5* for the case of a longitude invariant TEC. The calculated TEC is then in good agreement with the data. In order to test the stability of the algorithm, the *same set* of Doppler data was used with an added 10 % random noise. The results, shown in *fig. 6*, are found to be almost insensitive to the noise in the data.

Results are shown in *fig. 7* and *fig. 8* when the assumption of a longitude invariant TEC is no longer valid. It can be seen that although the calculated TEC can be in error of as much as 20 %, the general features of the data are still reproduced correctly.

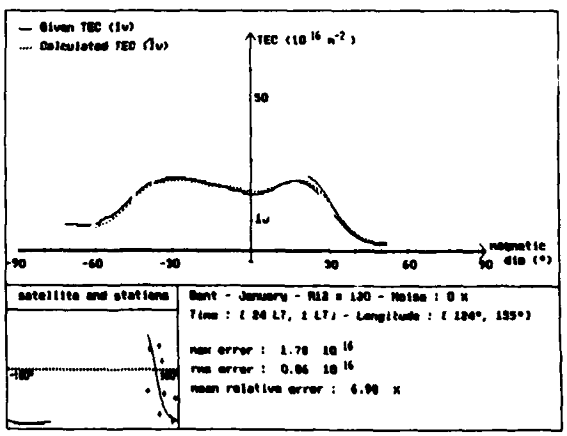


Fig. 5. Comparison of calculated TEC with simulation data. TEC independent of the longitude.

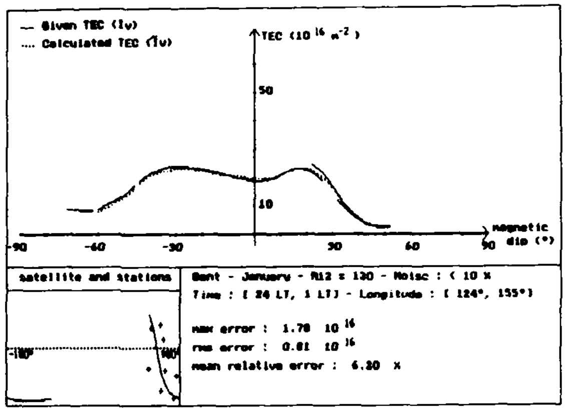


Fig. 6. Same as *fig. 5* with a 10 % random noise added to the Doppler data.

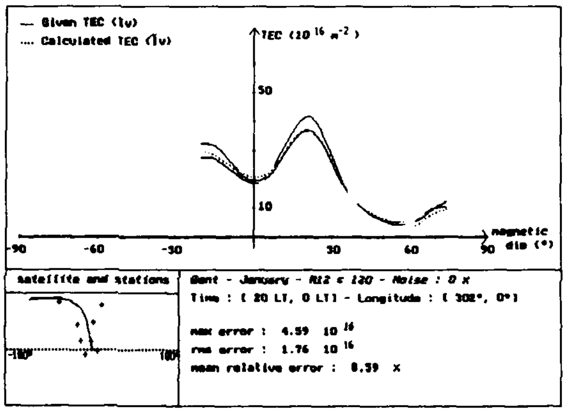


Fig. 7. Comparison of calculated TEC with simulation data. TEC slightly dependent on the longitude.

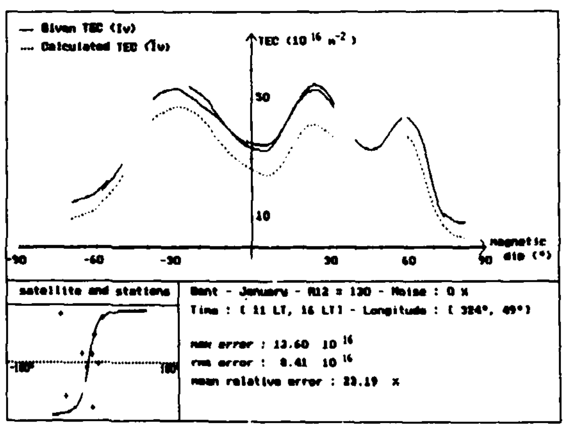


Fig. 8. Comparison of calculated TEC with simulation data. Combined effects of longitudinal and steep latitudinal TEC gradients.

4. MODEL CORRECTION OF IONOSPHERIC BIASES

4.1- Residual errors in a two-frequency system

Letting $f = f_u$ and $f = f_\ell$ in eq.(4) and solving with respect to L_0 yields :

$$L_0 = \hat{L}_0 + \lambda_2 + \lambda_3 + \lambda_R + O\left(\frac{1}{f^5}\right) \qquad (15)$$

with :

$$\left.\begin{aligned} \lambda_2 &= -\frac{\eta_u f_\ell - \eta_\ell f_u}{f_u f_\ell \left(f_u^2 - f_\ell^2\right)} \ell_2 \\ \lambda_3 &= \frac{1}{f_u^2\, f_\ell^2} \ell_3 \\ \lambda_R &= -\frac{f_u^2 R_u - f_\ell^2 R_\ell}{f_u^2 - f_\ell^2} \end{aligned}\right\} \qquad (16)$$

Typical values of λ_2, λ_3 and λ_R are shown in *fig. 1* for the DORIS 400/2000 MHz system. In the following, we will denote by λ_{3R} the sum of all terms of order ≥ 3 :

$$\lambda_{3R} = \lambda_3 + \lambda_R + O\left(\frac{1}{f^5}\right) \qquad (17)$$

In the next sections models for λ_2 and λ_{3R} as functions of the TEC are described.

4.2. Models for the 2nd and 3rd order terms

Starting from the expression of λ_2 given in eq.(4) :

$$\ell_2 = b \int_{\mathcal{L}} N_e B_0 \; |\cos\theta| d\ell$$

an altitude r can be found which satisfies the following equality :

$$\int_{\mathcal{L}} N_e B_0 |\cos\theta| d\ell = (B_0 |\cos\theta|)_r \int_{\mathcal{L}} N_e d\ell$$

or equivalently

$$\ell_2 = (B_0 |\cos\theta|)_r \frac{b}{a} \ell_1 \qquad (18)$$

r depends in general on N_e. However it can be shown by simulation - using a Chapman profile for N_e and the IGRF-1980 geomagnetic model - that eq.(18) holds with a good approximation for $r = r_m$, i.e. the altitude of the subionospheric point. By combining eqs.(16) and (18) a model for λ_2 is obtained whose output is :

$$\hat{\lambda}_2^m = -\frac{\eta_u f_\ell - \eta_\ell f_u}{f_u f_\ell \left(f_u^2 - f_\ell^2\right)} \frac{b}{a} (B_0 |\cos\theta|)_{r_m} \hat{\ell}_1 \qquad (19)$$

with $r_m = r_0 + 400$ km. The input of the model is $\hat{\ell}_1$ which is deduced from the TEC evaluations performed above.

A model for the term λ_{3R} can be derived by using a method similar to that proposed by *Clynch et al. (1979)*. Accordingly, a variable ρ is defined as :

$$\rho = f_u^4 \ \lambda_{3R}/\hat{\ell}_1^2 \tag{20}$$

A set of ρ values has been computed by simulating a satellite on a 1200 km circular orbit. The phase paths of signals propagating between the satellite and a ground station have been calculated by assuming a Chapman electron density profile with a scale height H given by :

$$H = \begin{cases} H_0 \text{ for } h \leq h_m F2 \\ H_0 + 25 \text{ km for } h > h_m F2 \end{cases}$$

The simulation cases were :

f_c = 5 ; 8 ; 11 ; 14 MHz
h_m = 250 ; 300 ; 350 km
H_0 = 40 ; 50 ; 60 km
E = 5 ; 10 ; 15 ... 90 °

with f_c the critical frequency of the ionosphere, h_m the height of the ionization maximum and E the elevation angle of the satellite at the ground station.

A scatter plot of the computed values of ρ as a function of the elevation angle shows that the data points are spread along a band of small width, making it possible to use a curve fitting method to model their variations. A polynomial regression algorithm was thus applied to the data. The resulting best fit was found to be :

$$\hat{\rho}^m(E) = 10^{-6} \sum_{k=0}^{k=5} c_k E^k \tag{21}$$

with : c_0 = 54.3136 ; c_1 = -2.3512 ; c_2 = 5.3717×10^{-2} ; c_3 = -5.1450×10^{-4} ; c_4 = 1.3141×10^{-6} ; c_5 = 4.8280×10^{-9}.

Hence, the 3rd order term model :

$$\hat{\lambda}_{3R}^m = \hat{\rho}^m(E) \ \hat{\ell}_1^2/f_u^4 \tag{22}$$

4.3. Evaluation of the model

By taking eq.(17) into account, eq.(15) can be rewritten in the form :

$$L_0 = \hat{L}_0 + \lambda_2 + \lambda_{3R} \tag{23}$$

The initial (uncorrected) range error in a two-frequency system is :

$$\lambda_1 = L_0 - \hat{L}_0 = \lambda_2 + \lambda_{3R} \tag{24}$$

The total range estimate resulting from model correction is :

$$\hat{L}_0^m = \hat{L}_0 + \hat{\lambda}_2^m + \hat{\lambda}_{3R}^m \tag{25}$$

Let us denote by λ_2' and λ_{3R}' the residual errors of the 2nd and 3rd order term models respectively :

$$\lambda_2' = \lambda_2 - \hat{\lambda}_2^m$$

$$\lambda_{3R}' = \lambda_{3R} - \hat{\lambda}_{3R}^m$$

The total residual range error is :

$$\lambda_1' = L_0 - \hat{L}_0 = \lambda_2' + \lambda_{3R}' \tag{26}$$

The evaluation of the correction model has been carried out by using simulation data. The results are shown in *fig. 9* for the DORIS system. The uncorrected range error can be as high as 12 cm at low elevation angles. The corrected range error has been plotted assuming various biases on the TEC estimate. The case corresponding to no TEC bias yields the ultimate limits of the correction model. In this case, the range error can be brought down to less than 1 cm.

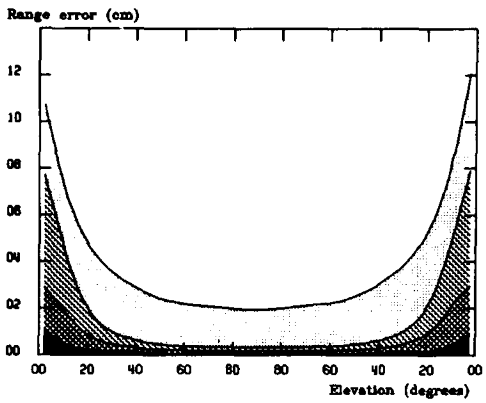


Fig. 9. Evaluation of the correction model for the 400/2000 MHz DORIS system. Range error vs. elevation angle of the satellite. Simulation data apply to a mid-latitude station. Highest curve (dotted area) shows the uncorrected error. Lower curves correspond to a systematic biais of 50 % (hatched area) and 20 % (cross-hatched area) in the TEC evaluation. The lowest curve (black area) is for a perfectly determined TEC and shows the limits of the model.

5. CONCLUSION

The accuracy of orbit determination systems based on dual-frequency Doppler measurements can be considerably increased by model correction of the 2nd and 3rd order ionospheric error terms. The magnitude of the residual corrected range error depends on the availability and accuracy of the TEC data for the satellite to ground station paths. Simulation results show that *sensible range error improvements* can be achieved even when the TEC is assumed to be imperfectly determined. When the TEC is perfectly known, the model can correct about 75 % of the ionospheric error. This gives, in the case of the DORIS 400/2000 MHz system, an ultimate bound to the ionospheric error of less than 1 cm.

Relevant TEC data can be derived from differential Doppler measurements on condition that a sufficiently dense network of ground stations exists so as to allow a certain amount of overlap between measurements from one station to the next. For sparse portion of the station network, such measurements may have to be complemented by TEC data from other sources. The need for complementary data (e.g. from *statistical TEC models or other systems*) and the way these data could be incorporated in the data bank still remains to be investigated in detail.

APPENDIX

Derivation of the vertical TEC in the vicinity of the satellite track

A distance can be defined in the space of the functions I_v by means of the norm :

$$\|I_v\|^2 = \sum_{i=1}^{p} \int_{\Theta_i} (I_{vi}^2 + I'^2_{vi})\, d\theta \qquad (A.1)$$

The *coefficients* $a_i(\theta)$ appearing in eq.(13) are derived from the Doppler measurements. They can be regarded as the components of a data-vector **a** :

$$\mathbf{a} = \begin{pmatrix} a_1 \\ \cdot \\ \cdot \\ \cdot \\ a_p \end{pmatrix} \qquad (A.2)$$

A norm can also be defined in the space of data-vectors by :

$$\|a\|^2 = \sum_{i=1}^{p} \int_{\Theta_i} (a_i^2 + a_i'^2)\, d\theta \qquad (A.3)$$

The scalar product associated to this norm will be denoted by <.,.>.

Since a variety of measurement and modelling errors have been neglected, I_v cannot be determined in an exact manner. Instead, we shall look for a function $\tilde{I}_v$ that approximates I_v. In order to keep the calculations tractable, we shall look for $\tilde{I}_v$ in a space of finite dimension. This can be done in the following way :

Let θ_k, ($k = 1,\ldots,n$) be the knots of a mesh defined on Θ in such a way that each Θ_i boundaries belong to the mesh *(fig. A1)*. The function $\tilde{I}_v$ will be required to be a 3rd order *derivable piecewise polynomial with a continuous derivative* :

$$\tilde{I}_v = \sum_{k=1}^{k=n} (x_{2k-1}\Phi_k + x_{2k}\Psi_k) \qquad (A.4)$$

where Φ_k and Ψ_k are basis functions in the space of such 3rd order piecewise polynomials on Θ *(fig. A2)*. Then, approximate data-vectors can be defined as :

$$\tilde{a} = \begin{pmatrix} \tilde{a}_1 \\ \cdot \\ \cdot \\ \cdot \\ \tilde{a}_p \end{pmatrix} \qquad (A.5)$$

with

$$\tilde{a}_i = \tilde{I}_v - b_i.c_i = \sum_{k=1}^{k=n} (x_{2k-1}\Phi_k + x_{2k}\Psi_k) - b_i c_i \qquad (A.6)$$

for $\theta \in \Theta_i$, $(i = 1,\ldots,p)$.

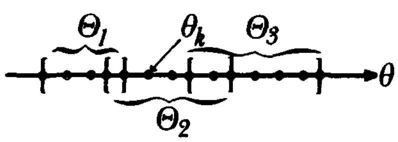


Fig. A1. Example of mesh defined on the latitude interval Θ.

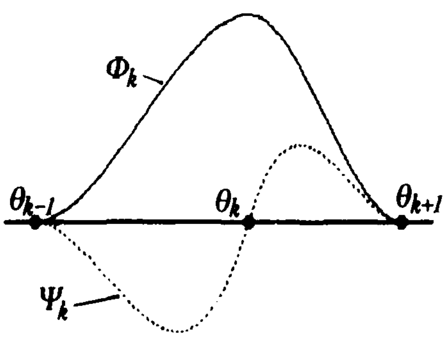


Fig. A2. Basis functions in the space of derivable 3rd order piecewise polynomials with continuous derivative.

Let $\tau_k = \dfrac{\theta - \theta_k}{\theta_{k+1} - \theta_k}$ then :

$$\Phi_k = \begin{cases} (\tau_{k-1})^2(3 - 2\tau_{k-1}) & \text{on } [\theta_{k-1}, \theta_k] \\ (1-\tau_k)^2(1+2\tau_k) & \text{on } [\theta_k, \theta_{k+1}] \\ 0 & \text{elsewhere} \end{cases}$$

$$\Psi_k = \begin{cases} -\tau_{k-1}^2(1-\tau_{k-1})(\theta_k - \theta_{k-1}) & \text{on } [\theta_{k-1}, \theta_k] \\ \tau_k(1-\tau_k)^2(\theta_{k+1} - \theta_k) & \text{on } [\theta_k, \theta_{k+1}] \\ 0 & \text{elsewhere} \end{cases}$$

Hence the basis functions of the data-vectors :

$$e_{2k-1} = \begin{pmatrix} \Phi_k \\ \cdot \\ \cdot \\ \cdot \\ \Phi_k \end{pmatrix} \; ; \; e_{2k} = \begin{pmatrix} \Psi_k \\ \cdot \\ \cdot \\ \cdot \\ \Psi_k \end{pmatrix} \; ; \; (k = 1,\dots,n)$$

$$e_{2n+k} = \begin{pmatrix} 0 \\ \cdot \\ \cdot \\ 0 \\ b_k \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \; ; \; (k = 1,\dots,p) \qquad (A.7)$$

Thus, any $\tilde{a}$ can be written as : $\tilde{a} = \sum_{k=1}^{k=m} x_k e_k$ *(A.8)*

with $m = 2n + p$. A necessary condition for $\tilde{I}_v \simeq I_v$ is that $\tilde{a} \simeq a$, or equivalently, $\|\tilde{a}-a\|$ min, i.e. $(\tilde{a}-a)$ should be perpendicular to every basis vector e_j :

$$\langle \tilde{a}-a, e_j \rangle = 0 \;, \quad (j = 1,\dots,m) \qquad (A.9)$$

The components x_k of $\tilde{a}$ are thus the solutions of a system of m equations with m unknowns :

$$\sum_{k=1}^{k=m} x_k \langle e_k, e_j \rangle = \langle a, e_j \rangle \;, \quad (j = 1,\dots,m) \qquad (A.10)$$

It can be shown that $\tilde{a} \to a$ when $n \to \infty$, i.e. when the mesh size is reduced. Then, $\tilde{I}_v \to I_v$ on condition that the intervals Θ_i are overlapping.

ACKNOWLEDGEMENT

This work was carried out with the support of CNES (Centre National d'Etudes Spatiales, Toulouse). The contribution of Pr. Ciavaldini (University of Rennes) for many helpful discussions is gratefully acknowledged.

REFERENCES

Budden K. G. (1966). Radio Waves in the Ionosphere. Cambridge Univ. Press.

Clynch J., Altenburg R. (1979). Ionospheric Residual Range Error Model. Proc. of the Second International Geodetic Symposium on Satellite Doppler Positioning, Austin, USA.

Lassudrie-Duchesne P., Fleury R., Bertel L., Monier L. (1986). Ionospheric Doppler Correction Model for Satellite Orbit Determination. Adv. Space Res., Vol. 6, No. 9,

Lassudrie-Duchesne P. (1986). A new approach to the derivation of the total electron content using Doppler measurements from a single ground station. International Beacon Satellite Symposium (URSI-COSPAR), pp. 121-134. Univ. Oulu, Finland.

Leitinger R., Schmidt G., Taurlainen A. (1975). An Evaluation Method Combining the Differential Doppler Measurements from Two Stations that Enables the Calculation of the Electron Content of the Ionosphere. J. Geophys. 41, 201-213.

Llewellyn S.K., Bent R.B. (1973). Documentation and description of the Bent ionospheric model. Rep. AFCRL-TR-73-0657.

DISCUSSION

L. BOSSY, BEL
English Translation

Evaluation of the TEC enables an approximate allowance to be made for the f^{-3} term in the Doppler effect. For downstream terms, the physical parameters required must take into account the shape of the profile, which is not the case with the TEC. What is your opinion?

AUTHOR'S REPLY
English Translation

It is true that the term 13R does not only depend on the CET but also, to a lesser extent, on the profiles of the electronic concentration crossed by the electromagnetic ray. Examination of the various components of 13R shows that we can expect the term R to be more sensitive to the shape of the profile. This was clearly confirmed by the simulations we made using different profiles all from the same CET.

Figure 1 shows that the term R (or λR) is generally low, except possibly at low angles of elevation. A the angles of elevation which interest us (i.e. $E > 10^\circ$), the influence of the profile on the term 13R (or λ3R) is negligible.

THE USE OF BROADCAST SIGNALS FOR PASSIVE SENSING IN AUTOMATED HF COMMUNICATION SYSTEMS

by

N.G.Riley
Hull-Warwick Communications Research Group
University of Hull
Hull HU6 7RX
UK

Abstract

The use of passive sensing for derivation of HF system control data has advantages over active techniques from both tactical and spectrum conservation points of view. In this paper methods are described by which ionospheric data may be derived from passive sensing of the HF spectrum. Data derived using these techniques may be used in the operational decision process to control automated hf communication systems. The techniques described are facilitated by the availability of powerful, low-cost signal processors and fast tuning receivers allowing a high level of signal processing to be performed in real time.

1. Introduction

The purpose of this paper is to present a number of concepts relating to the general topic of passive sensing of the HF spectrum as an aid to real time control of HF radio systems. The concepts and work described represent one of a number of themes being pursued by the Hull-Warwick Communications Research Group based at the Universities of Hull and Warwick.

The feasibility of implementing the techniques described has been dramatically increased by the current availability of cheap and powerful digital signal processors - HF signals and noise may now be analysed in real time in both the time and frequency domains using equipment which does not increase the overall cost of the communication system significantly.

The paper describes work on three aspects of passive sensing:

a) The design and development of a new HF propagation monitoring system which may be used for both experimental and operational purposes.

b) The development and comparative testing of a low cost 'chirpsounder' receiver.

c) The extraction of system control data from passive sensing of broadcast transmissions and HF interference/noise.

2. Channel Evaluation Methods

Over the past five years automated HF systems have developed from the primarily experimental stage [Elvy 1985] to the point where several companies are offering complete communication systems (Eg Harris 'AUTOLINK' Series RF-7100, Andrews Corporation CQS) in addition to those systems developed for military use (Eg Plessey AiCORN system, Marconi ASSATS system).

A vital component of any automated HF system is the frequency selection sub-system. Careful frequency selection is necessary to maintain communication on frequencies which are both supported by ionospheric refraction over the desired path and are free from interference and give adequate signal to noise ratio (SNR).

Automated HF communication systems commonly use propagation modelling and real time channel evaluation [Darnell 1983] to fulfil these needs.

Propagation models may range from simple empirical models such as MINIMUF-3.5 [Rose & Martin 1978] to more complex models [Davy et al 1987, Dick 1987, CCIR Supplement to Report 252] based on the use of the full CCIR numerical maps of the ionosphere [CCIR Report 340]. These more complex models often require the use of a 386 processor-based personal computer for their implementation. In some systems [Reilly & Daehler 1985 for example] the model may be updated using current ionospheric sounder measurements to improve accuracy. These model updates may be disseminated periodically over an additional HF system (Eg Prophet [Sailors 1984]).

In order to put into perspective the use of passive sensing of the HF spectrum for control of automated systems, it is instructive to consider the various types of real time channel evaluation (RTCE) systems currently in use:

Active RTCE:

Signals are radiated specifically for reception and analysis by the system to determine the channel capacity of a number of candidate channels. These radiated signals include various types of oblique ionospheric soundings (Barry Research Chirpsounder [Barry & Fenwick 1975]), Channel Evaluation and Calling (CHEC) [Stevens 1968], pilot tone systems (Andrews Corporation, CQS [Parkins & McNamara 1989]) and systems which simultaneously radiate on several channels (Eg Marconi ASSATS).

Link Quality Measures or Passive RTCE

Here, channel quality data is derived from measurements made of some parameter of the received operational traffic. The most useful indicators of channel quality are those parameters closely related to the way that the message signal is coded, such as signal to noise ratio, zero-crossing analysis [Shaw et al 1988] or phase error measurement on DPSK transmissions [Riley 1987].

These techniques are refered to as Passive RTCE since they do not require radiation of probing signals and hence help to minimise congestion of the HF spectrum.

Noise Assessment

A further type of measurement, closely related to Passive RTCE consists of determining the relative channel capacity of a number of channels by assessing the level of background noise and interference on each channel. Measurements may be performed on a set of channels which are allocated to the system but not currently in use. Several systems (eg Andrews RTFM and [Elvy 1985]) employ channel selection schemes based on a ranking of channels derived from both current and historical measurements.

This paper describes progress and work being carried out at the Universities of Hull and Warwick on several aspects of a further class of measurements, where general propagation data may be derived from passive sensing of the HF spectrum and applied in a number of ways to the control of automated HF systems.

The work primarily centres around the use of measurements of broadcast signals of various types but also includes an investigation of the usefulness of measurements of noise and interference.

3. Enabling Technology for work on Passive Sensing and RTCE

The current interest in passive sensing techniques is driven by the recent availability of cheap, fast tuning computer controllable receivers, powerful personal computers and low-cost signal processors. This equipment is now available at a cost which is a small fraction of the cost of the radio system itself and therefore becomes commercially attractive. Use of this technology is central to current work at Hull and Warwick on efficient coding and multiple-access methods and modem design in addition to the work described in this paper.

The particular equipment standardised for most of these studies is as follows:

Receivers/Transceivers:	ICOM ICR71E / IC735
Computer:	Various 286- and 386-based PC compatibles
Digital Signal Processors:	Texas Instruments TMS320-C25 with Loughborough Sound Images (LSI) development system
Stand-alone processors:	STE rack system

The following sections outline three areas of study: Propagation Monitoring System, Low-cost Oblique Sounder Receiver and derivation of system control data from HF spectrum monitoring.

4. HF Propagation Monitoring System

The Propagation Monitoring System [Darnell et al 1988, Darnell & Hague 1989] was developed for CCIR Study Group 6 with support from the UK Department of Trade and Industry. A prototype using only one transmitter was successfully demonstrated at the Study Group 6 meeting in Geneva, 2-6 May 1988.

The system, when fully implemented, will consist of a number of broadcast transmitters worldwide which will radiate a specially designed signal format on several frequencies in a given sequence.

The format of the signal, which is repeated every 12 seconds, contains elements which allow propagation information to be derived from the received signal at various levels of sophistication. The signal format consists of the following (See Figure 1):

FSK Preamble (1 Second) 100 bits/sec

Allows initialisation of receiver AGC and provides a degree of tolerance to slight timing errors throughout the system.

CW Identification Sequence (3.3 Seconds)

The higher of the FSK tones is keyed, for example with morse code, to provide identification capable of audio interpretation.

Complementary Sequences (0.53 Seconds)

Two 256-bit binary complementary sequences are transmitted at a rate of 1200 bits/sec, each followed by an interval of 50 ms to allow for recovery from multipath propagation. Again the upper FSK tone is used. Matched filtering at the receiver allows an estimate to be made of the channel impulse response and therefore multipath structure.

FSK Reversals (4 Seconds) 100 bits/sec

This section allows characterisation of short-term fading (by use of a Law-assessor FSK demodulator), estimation of SNR by comparing levels in the mark and space intervals and characterisation of noise/interference by Fast Fourier Transform analysis of the signal received in the space intervals.

CW Signal (3 Seconds)

A further CW signal is transmitted to bring the total signal time to twelve seconds, after which a frequency change takes place.

In a system operating with a number of transmitters, a frequency schedule is designed so that each transmitter transmits continuous repetitions of the signal for a number of minutes before changing to its next frequency. During this time the receiver will scan all transmitters several times, depending on the number of transmitters. Figure 2 shows one possible schedule.

A data reduction system has been designed which dumps reduced data to floppy discs for subsequent analysis at a central site.

The system is intended to be available both for use by the CCIR to aid production of a world HF propagation database, and for use by individuals or organisations for experimental purposes or real time system management.

Design information for compatible terminals is now available. In addition, a fully engineered version of the system will shortly be tested.

5. Development of a Low-cost Chirpsounder Receiver

HF propagation data may be derived for a system by reception of chirpsounder transmissions which are radiated from more than 40 sites throughout the world. Given that the transmissions already exist, this technique may be considered in some sense as a passive monitoring technique.

Since it is often not cost-effective to incorporate a Barry Research Chirpsounder Receiver into a small HF system, a cheaper alternative has been developed [Jowett et al 1989] using the equipment described in section 3. The system has been shown to give a reliable estimate of maximum usable frequency (MUF) but currently is not able to deduce the multipath structure of a link.

The chirpsounder detector consists of a matched filter and a threshold detector implemented on the TMS320-C25 processor. The matched filter uses a sample rate of 8 kHz and 240 filter taps. The system architecture is shown is Figure 3.

Prior to following a chirp transmission from a particular transmitter, the system performs an assessment of interference on channels at 200 kHz intervals throughout the HF band. An attempt is made to replace any channels suffering high levels of interference or noise with channels 10 kHz and 20 kHz higher in frequency. If neither of these additional channels are interference-free then the frequency step is eliminated from the list of frequencies to be scanned for received chirp signals.

At the correct time, the system tunes ahead of the chirp sweep to each of the frequency steps in turn, recording whether or not a chirp is detected on each channel. The output is given in the form shown in Figure 4, from which an estimate of the MUF on the path between the transmitter and receiver may be made by either manual or automatic means.

Trials have been carried out to compare the output from this system with a standard Barry Research Chirpsounder Receiver with encouraging results. A typical example of results from a back-to-back trial is shown in Figure 5. The upper part of Figure 5 shows an oblique ionogram from the Barry Receiver for a path from Milltown, Scotland to Farnborough, England. The lower half shows the output from the Hull chirp detector, where solid bars indicate received chirp signals. In both cases the MUF may be estimated at around 15 MHz with intermittent returns from sporadic E at frequencies up to 24 MHz.

Future development of the system could include the use of analogue filter taps, use of a wider receiver bandwidth (3 kHz is used at present) with compensation for receiver characteristics and resolution of multipath structure.

6. Extraction of HF System Control Data from Passive Sensing of Broadcast Transmissions and HF Noise and Interference

One technique used by HF radio operators to assess the range of frequencies which will propagate over a particular path is to tune in to known broadcast signals from transmitters in the vicinity of the far end of the desired link. The technique relies on the skill and experience of the operator in chosing signals which may be readily identified and whose source location is known.

For use to be made of this technique in an automated system, these acknowledged problems of positive signal and transmitter site identification must be overcome. One way of at least partially overcoming the problems is to try to impart to the system the knowledge and experience of the operator ie. to construct an expert system.

The first step in such a process is to identify classes of signals which may be useful for passive sensing purposes. These classes of signals may include:

a) Broadcast service transmissions.

b) Oblique sounder signals and signals radiated for propagation analysis purposes.

c) Standard time signals, beacons (eg Radio Amateur) and experimental broadcasts.

d) The HF occupancy spectrum

Signals in class b) above were discussed in Sections 4 and 5 of this paper. Some ideas on the use of signals in classes a), c) and d) are presented below.

6.1 Broadcast Service Transmissions

Several techniques may be employed to try to overcome the problem of signal identification. the simplest and least sophisticated is to chose "reliable" broadcast transmissions whose source location is known and which do not generally suffer from co-channel interference at the receiver site.

An example of the use of such a carefully-chosen broadcast service is the reception in the UK of broadcasts from Swiss Radio International, Berne on a suite of frequencies including 3985, 6165, 9535 and 12030 kHz. When scanned sequentially these transmissions can provide MUF data in a form shown schematically in Figure 6. Short-term ionospheric forecasting agencies such as those run by FTZ in Germany and Marconi Research in the UK monitor several such frequencies.

Other techniques for signal identification which require a higher degree of signal processing may include speech/pitch detection to identify signals above noise and recognition of identification announcements and interval signals or "jingles", whose transmission schedules are known. These techniques have yet to be investigated but it is anticipated that, using an expert system approach allied to a broadcast station database, some progress may be made.

6.2 Noise Analysis

Many systems use an analysis of noise and interference on specific channels as part of a frequency selection procedure (See section 2 above). In this study a more general analysis is investigated where the spectrum of HF noise and interference is analysed in a statistical way to derive general propagation information.

The noise and interference on which any analysis would be performed is that arriving at the receiver over the whole HF spectrum at a particular instant. The noise and interference received in a given channel generally consists of signals from a variety of transmitters at various ranges from which propagation is possible, in addition to both manmade and natural radio noise propagated over similar paths. It may therefore be expected that the area from which propagation is possible will vary both in size and location with changes in the state of the ionosphere, characterised in particular by changes in the F2-layer critical frequency f_oF2. Statistically, therefore, the amount of noise and interference received may also be expected to vary with frequency relative to f_oF2 at some point along the path.

The aim of the current study is to identify what, if any, general propagation data may be derived in real time from a consideration of this noise and interference spectrum.

Several databases of spectral occupancy have been built up in recent years [Gott et al 1982, Wilkinson 1982 for example], of which that due to Gott and co-workers has relevance to conditions in the UK. Analyses of this database have shown several features worthy of note [Wong et al 1985, Riley 1986]:

a) The average power received varies from one user band to another, as expected due to the different average levels of power radiated.

b) The occupancy falls off statistically at a frequency which is believed to be related to the MUF for a 4000 km path.

It is recognised that hf spectral occupancy varies considerably between, for example, the UK and the USA. Conclusions drawn from an analysis of spectral occupancy in Europe should therefore be used with caution elsewhere.

Any analysis of received noise / interference power as a function of frequency must take account of the differences in typical transmitter power in the various user bands. Figure 7 shows the median received occupancy as a function of frequency for channels in the 'Fixed' and 'Fixed and Mobile' bands for various times of day. A shift in the high frequency fall-off is apparent.

Further investigation of this possible relationship will be pursued using a computer-controlled spectrum monitor designed for the purpose.

6.3 Standard Time Signals and Beacons

Time signal transmissions and hf beacons may be used in a similar manner to broadcast services except that they are in general easier to identify and the transmitter position is known - the identification problem reduces to that of discriminating against noise and interference.

There are several well-known time signal transmissions of which the following are examples:

USA:	WWV	5, 10, 15, 20 MHz (Male Voice)
	WWVH	5, 10, 15 MHz (Female Voice)
Germany:	NAUEN	44.525 MHz
Hong Kong:		4.2, 8.5, 13.0, 17.1, 22.5 MHz
Spain:		6.840, 12.008 MHz
USSR:		4.996, 9.996 14.996 5.000, 10.000, 15.000 5.004, 10.004, 15.004 MHz

WWV and NAUEN are continuous transmissions whereas some of the other signals are only transmitted at set

times. For example the Spanish signal is radiated on 12 MHz from 10.00 to 10.25 and on 6.8 MHz from 10.30 to 10.55. In addition the digital transmissions such as NAUEN are easier to detect automatically. Again this monitoring technique is expected to yield "snippets" of propagation information which may be incorporated into an overall picture of prevailing propagation conditions, for example by the use of artificial intelligence techniques.

HF beacons provide another source of propagation information, including the possibility of assessing the lowset usable frequency using beacons which transmit at various power levels according to a predetermined schedule.

7. Summary

Work being undertaken on passive sensing of broadcast signals for control of automated hf systems has been described, including development of two specific systems and ideas for passive sensing of general broadcast signals and "noise". The low-cost chirpsounder receiver and propagation monitoring system have both undergone successful trials whilst work on passive sensing of general broadcast signals is at an early stage.

The techniques described involve a high level of processing of the received signals in real time and have been facilitated by the current availability of cheap, powerful signal processors and development packages and low-cost fast tuning receivers.

The paper has concentrated only on the derivation of propagation data from the hf spectrum - the way in which a system is controlled using this data also forms a major element of the current study.

8. References

Barry, G.H. and R.B. Fenwick, 1975. "Techniques for real-time hf channel measurement and optimum data transmission", AGARD CP-173 "Radio systems and the ionosphere", Athens.

Darnell, M., 1983. "Real time channel evaluation", AGARD Lecture Series No. 127 on "Modern HF Communications".

Darnell, M. and J. Hague, 1989. "Design and performance of an hf real-time channel evaluation (RTCE) system", IEE Colloquium "Adaptive HF Management", Digest No. 1989/33.

Darnell, M., J. Hague and A. Chan, 1988. "A global common-user system for the provision of hf propagation data", AGARD EPP Symposium on "Propagation effects and circuit performance of modern military radio systems with particular emphasis on those em·'oying band spreading", Paris.

Davy, P., R. Hanbaba, M. Lissillour and H. Sizun, 1987. "Microcomputer based methods for the estimation of hf radio propagation and circuit performance", Proc. IEE Fifth International Conference on Antennas and Propagation, York, IEE CP 274.

Dick, M.I. and B.H. Miller, 1987. "Microcomputer - based method for the estimation of hf radio - circuit performance", Proc. IEE Fifth International Conference on Antennas and Propagation, York, IEE CP 274.

Elvy, S.J., 1985. "A design for an automatic hf radio system", Radio Science, Vol. 20, No. 3, pp261-268.

Gott, G.F., N.F. Wong and S. Dutta, 1982. "Occupancy measurements across the entire hf spectrum", AGARD Conference Proceedings 322, "Propagation aspects of frequency sharing, interference and systems diversity".

Jowett, A.P., M. Darnell and N.G. Riley, 1989. "Passive monitoring of chirpsounder transmitters as an aid to hf frequency management", IEE Colloquium "Adaptive Frequency Management" Digest No. 1989/33.

Parkins, W.A. and L.F. McNamara, 1989. "Practical enhancements for hf systems", IEE Colloquium "Adaptive HF Management", Digest No. 1989/33.

Reilly, M.H. and M. Daehler, 1985. "Sounder - updated statistical model predictions of maximum usable frequency for hf skywave predictions", NRL Memorandum Report 5677, Naval Research Laboratory, Washington DC.

Riley, N.G., 1986. "An hf interference model - Early steps", IEE Colloquium on "HF frequency management", London, Nov 1986.

Riley, N.G., 1987. "Real time channel evaluation for a 2.4 kbps parallel tone modem", presented at the IEE Fifth International Conference on Antennas and Propagation, York.

Rose, R.B. and J.N. Martin, 1978. "MINIMUF - 3.5", NOSC TD 201, AD-A066-256.

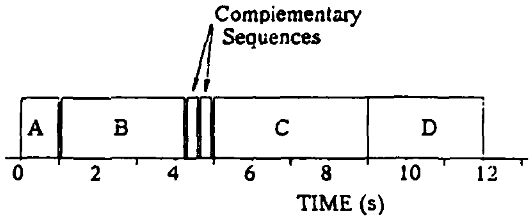
Sailors, D.B., 1984. "Tactical decision aids for hf communication", NOSC TD 782, AD-B091-227.

Shaw, M., B. Honary and M. Darnell, 1988. "An RTCE-assisted ARQ Transmission Scheme: Design and Performance", Proc. IEE Fourth International Conference on HF Radio Systems and Techniques, London, IEE CP 284.

Stevens, E.E., 1968. "The CHEC sounding system", in "Ionospheric radio communications", Ed. K Folkstad, Plenum Press, pp359-369, New York.

Wilkinson, R.G., 1982. "A statistical analysis of hf radio interference and its application to communications systems", IEE International Conference on HF Communication Systems and Techniques, IEE CP 206.

Wong, N.F., G.F. Gott and L.W. Barclay, 1985. "HF spectral occupancy and frequency planning", Proc. IEE, Vol 132, Part F, No.7.



A - FSK

B - CW Identification Sequence

C - FSK Reversals

D - Single Tone

Figure 1: Signal Format for Propagation Monitoring System

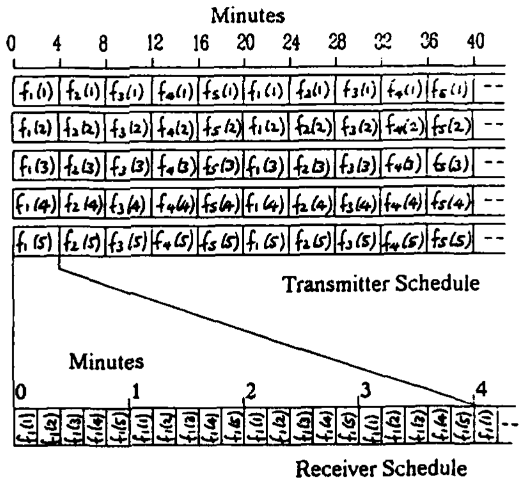


Figure 2: Frequency Schedules for Propagation Monitoring System

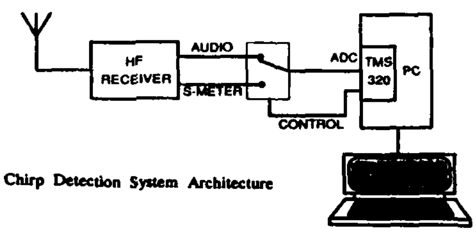


Figure 3: Chirp Detection System Architecture

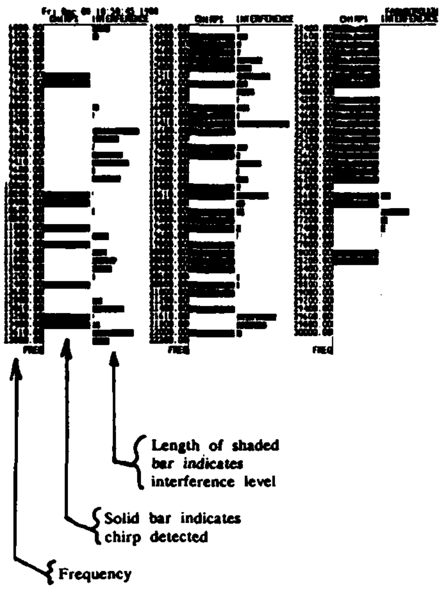


Figure 4: Chirp Detector Output

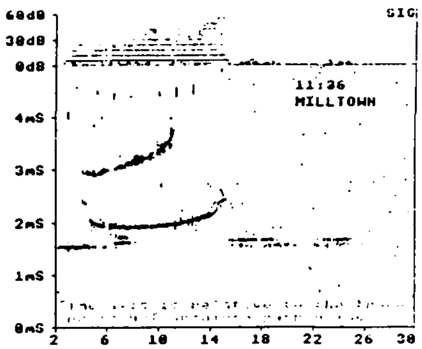


Figure 5a: Oblique Ionogram

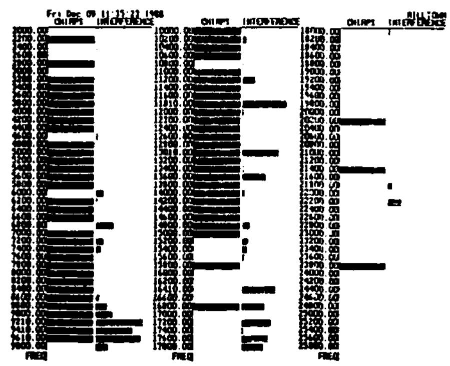
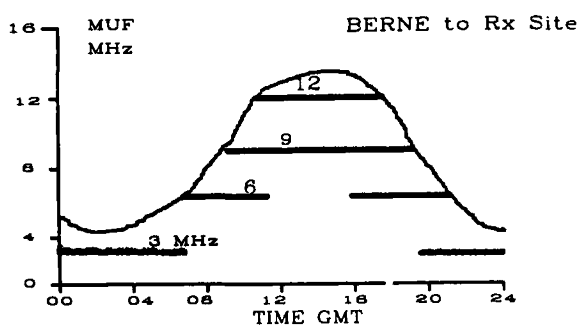


Figure 5b: Corresponding Chirp Dectector Output

Figure 6:
MUF and Propagation Duration for a Suite of Broadcast Frequencies



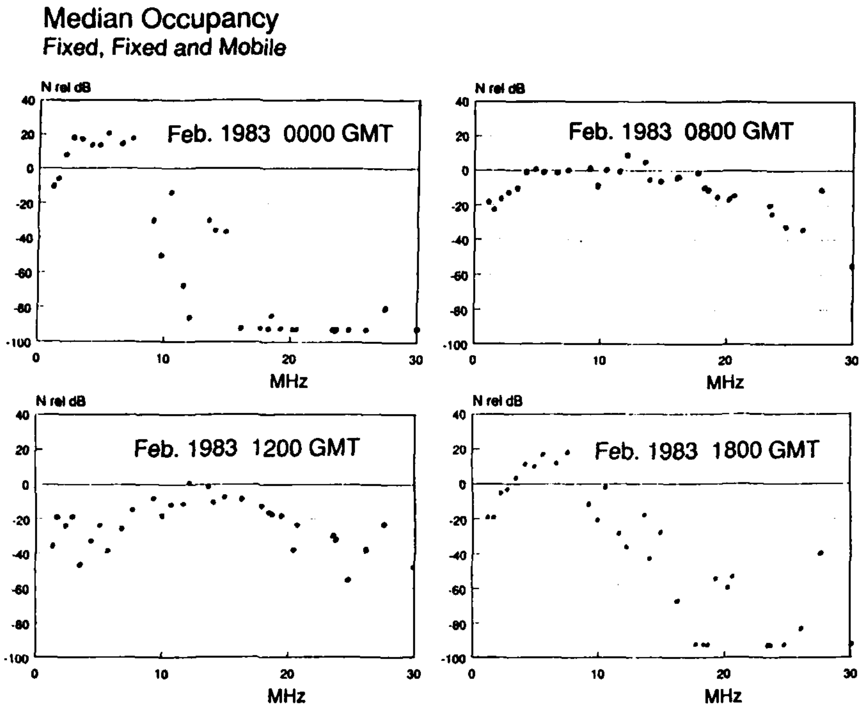


Figure 7: Median Occupancy at Various Times of Day

DISCUSSION

J. BELROSE, CAN

Isn't it true that a Chirpsounder is neither passive nor strictly evaluates a communications channel?

AUTHOR'S REPLY

I agree that the Chirpsounder system does not strictly evaluate a communications channel. However, it is of use in channel selection and it is not a passive system. Our aim is to develop a system which is appropriate to a low-cost HF communication system and which uses Chirp transmissions in the context of "signals of opportunity."

PREVISIONS A TRES COURT TERME
PAR MODELISATION DU BROUILLAGE ET DE L'IONOSPHERE
A L'AIDE DE LA RETRODIFFUSION

J. CARATORI et C. GOUTELARD
L E T T I - UNIVERSITE PARIS-SUD
9, av. Division Leclerc B.P. 16 - 94231 CACHAN CEDEX - FRANCE

ABSTRACT -

H.F radio links involving reflexion from the ionosphere are mainly long-range communications.

The setting up of an optimum link, requires the knowledge of the real propagation and interference data. Several types of ionospheric predictions, based on different ionospheric soundings, are necessary in order to take account of the medium variability.

A centralized backscatter sounding station is well adapted to provide short-term predictions. However, the interpretation of the data, and the inversion of the ionogram are difficult. A method of inversion, taking account of horizontal gradients is proposed. Experiments carried out at the station STUDIO of the LETTI showed its validity.

Interferences are very important in the H.F. band, and represent the second factor to deal with. The spatial decorrelation of interferences is highlighted, and the derivation of predictions for the Western Europe zone is discussed. These predictions are then used to compute the signal to noise ratio of a link.

I. - INTRODUCTION -

Les télécommunications en ondes décamétriques par réflexion dans l'ionosphère sont essentiellement réservées aux liaisons longues distances.

l'établissement de liaisons optimales doit tenir compte à la fois des conditions réelles de propagation et de brouillage. La variabilité du milieu nécessite l'établissement de prévisions ionosphériques élaborées à partir de différents types de mesure bien connus. L'utilisation de la méthode de sondage par rétrodiffusion offre l'avantage d'une station unique, et permet l'établissement de prévisions à très court terme.

Le problème de l'inversion de l'ionogramme est cependant difficile à résoudre. Une méthode, prenant en compte les gradients d'ionisation horizontaux, est présentée. Son expérimentation, effectuée à la station STUDIO du LETTI, en a montré la validité.

Le brouillage, très important dans cette gamme de fréquence est le second facteur à prendre en considération. Sa décorrélation spatiale est mise en évidence, et l'établissement de prévisions dans la zone Europe occidentale est abcrdée.

La prévision du rapport signal à bruit d'une liaison peut alors être effectuée.

II. - MODELISATION DE L'IONOSPHERE -

II.1. Introduction.

L'ionosphère est un milieu éminemment variable en fonction du lieu et du temps. Ses caractéristiques sont connues grâce aux mesures horaires effectuées en une centaine de points du globe, là où se trouvent des stations de sondage zénithal [1]. Cette manière de procéder est suffisante pour des études de morphologie globale du milieu, comme pour des études statistiques, mais est inadaptée à la gestion en temps réel des réseaux de communciation. En effet, toutes les informations collectées ne sont pas immédiatement disponibles et, même si elles l'étaient, l'utilisateur serait conduit à les interpoler de façon plus ou moins rigoureuse, particulièrement en présence de perturbations, du fait du sous-échantillonnage spatial et temporel des mesures [2].

Les erreurs de mesure pourraient être réduites en augmentant le nombre de sondeurs ainsi que la périodicité des sondages, mais un tel système serait à la fois coûteux et compliqué, car il nécessiterait la collecte, le traitement et la distribution rapide de données, sur une grande échelle.

La plupart des problèmes évoqués sont théoriquement résolus grâce à une station centralisée de sondage par rétrodiffusion équipée d'une antenne orientable tous azimuts. En effet, une telle station est à même de donner en temps réel une vue panoramique des conditions de propagation, dans une zone dont le rayon peut dépasser 3000 km.

Bien que très séduisante dans son principe, cette méthode n'a connu jusqu'à présent qu'un faible développement, car les mesures par rétrodiffusion sont beaucoup plus difficiles à interpréter que les mesures zénithales. En particulier l'opération fondamentale d'inversion, qui consiste à déduire le profil d'ionisation de l'ionogramme de rétrodiffusion, est beaucoup plus compliquée qu'en zénithal.

Ce problème a été étudié depuis 1967, notamment par HATTFIELD aux Etats-Unis, et par GOUTELARD en France, et déjà ces auteurs avaient suggéré de stabiliser la solution par la mesure de l'angle d'élévation.

Plus tard, en 1974, divers auteurs, comme RAO et YEH, ont proposé des méthodes d'inversion rigoureuses n'utilisant que la trace frontale de l'ionogramme. Ces méthodes donnaient de bons résultats lorsqu'elles étaient appliquées à des ionogrammes simulés, mais divergeaient lorsqu'elles étaient appliquées à des ionogrammes réels.

En 1977 GOUTELARD et moi-même avons proposé de stabiliser la solution du problème inverse en utilisant, non pas des points localisés de l'ionogramme, mais la totalité des points disponibles, et en faisant appel de surcroît à des informations supplémentaires, telles que celles fournies par des sondages zénithaux par exemple.

Plus tard, nous avons proposé d'employer en plus des sondages par rétrodiffusion classiques, des sondages à azimut variable, solution rendue possible grâce à la station de sondage par rétrodiffusion STUDIO du LETTI. Plusieurs articles ont déjà été publiés sur ce sujet.

La méthode originale présentée aujourd'hui, est l'une des méthodes d'inversion que nous avons développées. Elle prend en compte des gradients d'ionisation horizontaux, et fait appel aux sondages à azimut variable.

II.2. Problèmes posés par l'inversion de l'ionogramme de rétrodiffusion.

Considérons le problème d'un point de vue très général. A un instant donné, le profil d'ionisation est défini par une fonction de la forme :

$$N = f(h, Az, d) \qquad f \in \mathfrak{L}$$

avec :

- N concentration électronique (m^{-3})
- h altitude
- Az azimut du sondage
- d distance par rapport au sondeur
- $\mathfrak{L}$ ensemble des profils d'ionisation possibles.

De même, si nous postulons que seule la trace frontale de l'ionogramme de rétrodiffusion est exploitable (car bien définie), nous pouvons décrire celle-ci par une équation de la forme :

$$T_F = g(F, Az) \qquad g \in \mathfrak{J}$$

avec :

- T_F temps de focalisation (minimum du temps de ppropagation de groupe)
- F fréquence d'émission
- $\mathfrak{J}$ ensemble des ionogrammes possibles (calculés et mesurés).

Il existe une relation entre les fonctions f et g, symbolisée par un opérateur A tel que :

$$g = A(f)$$

Calculer g connaissant f, c'est résoudre le "problème direct", c'est-à-dire simuler l'ionogramme de rétrodiffusion à partir du profil d'ionisation. Résoudre le "problème inverse", c'est inverser l'opérateur A, c'est-à-dire exprimer f en fonction de g :

$$f = A^{-1}(g)$$

En fait, cette notation simple cache un problème très complexe, car l'opérateur A jouit des propriétés suivantes :

- il n'a pas d'expression analytique
- il est non linéaire
- il est non bijectif
- il n'est pas toujours continu.

Il en résulte que la solution peut ne pas exister, ou ne pas être unique et que, de plus, le problème inverse n'est pas stable. L'instabilité se manisteste, en pratique, par le fait qu'une très faible variation des données, due aux erreurs de mesure (bruit d'appareillage, bruit de troncature, bruit de quantification), entraîne une grande variation de la solution. Dans l'état actuel de nos connaissances scientifiques et technologiques, il paraît donc illusoire d'espérer obtenir une solution rigoureuse.

II.3. Méthodes actuelles d'inversion de l'ionogramme de rétrodiffusion.

L'inversion de l'ionogramme de rétrodiffusion n'est possible que grâce à une redéfinition du concept de solution. Le plus souvent, on a recours aux "méthodes d'essais" et de "quasi-solution", qui consistent à résoudre une fois pour toutes le problème direct, et à adopter comme solution approchée (ou solution généralisée), la fonction $\tilde{f}$ telle que :

$$||A(\tilde{f}), g|| = \text{Inf } ||A(f), g|| \qquad f \in \mathfrak{Z}$$

Lorsque cette méthode est en défaut, c'est le cas lorsque $\mathfrak{Z}$ n'est pas compact, on utilise plutôt la "méthode de régularisation" [3] [4].

Quelle que soit la méthode employée, le calcul effectif de la solution se ramène à un problème de recherche du minimum lié d'une fonction de plusieurs variables ; c'est un problème d'optimisation sous contraintes. Une multitude d'études ont été publiées sur la question de l'inversion de l'ionogramme de rétrodiffusion ces 20 dernières années. A une exception près, seules les méthodes d'essais et de quasi-solution ont été utilisées, car elles sont d'apparence simples à développer et à mettre en oeuvre, et ne nécessitent, en général, qu'un investissement théorique réduit [5] [6] [7] [8] [9]. Elles font toujours appel à un processus itératif et, dans la majorité des cas, apparaissent comme des méthodes de correction. Le profil initial, ou "profil d'ordre zéro", est généralement issu des prévisions à long terme, ou bien d'un sondage zénithal.

Malgré tous les efforts déployés pour tenter d'inverser l'ionogramme de rétrodiffusion, aucune solution totalement satisfaisante ne semble avoir été proposée à ce jour. La plupart des méthodes présentées supposent que la trace frontale est connue avec une très grande précision [5] [7] [8], ce qui n'est malheureusement jamais le cas en pratique, ne serait-ce qu'à cause des incertitudes de mesure qui peuvent entraîner l'instabilité du processus d'inversion ou, dans le meilleur des cas, conduire à unne précision globale rédhibitoire. Ces remarques nous ont amené à reconsidérer le problème de l'inversion, processus qui a abouti au développement de méthodes originales [10], dont la méthode qui va être présentée ci-après.

II.4. Elaboration d'une nouvelle méthode d'inversion de l'ionogramme de rétrodiffusion.

La méthode proposée ici part de la remarque selon laquelle le profil exact est, dans l'état actuel de nos connaissances, impossible à obtenir par inversion de l'ionogramme de rétrodiffusion. Nous nous sommes donc fixé dès le départ comme objectif de déterminer, non pas un profil d'ionisation exact, mais un profil d'ionisation moyen valable dans une zone de 2000 à 3000Km de rayon autour du sondeur, et celà en temps réel. Il est clair que la connaissance de ce profil moyen constituerait déjà une avancée considérable, surtout si les moyens techniques mis en oeuvre restaient limités.

La solution à laquelle nous avons abouti repose sur une double modélisation, celle du profil d'ionisation, et celle de la trace frontale de l'ionogramme de rétrodiffusion. Cette technique offre l'avantage de réduire chaque modèle à un petit nombre de paramètres, entre lesquels il est possible d'établir des relations numériques.

Notons que le fait de remplacer la trace frontale de l'ionogramme par un petit nombre de paramètres, équivaut à un moyennage, ou à un lissage, des mesures. Il en résulte que l'incertitude affectant chaque paramètre est bien inférieure à celle affectant les données initiales, ce qui contribue à stabiliser le processus d'inversion.

a - Modélisation de l'ionosphère -

Le modèle utilisé est un "modèle d'ordre un", ou "modèle différentiel", simplifié, obtenu en introduisant des gradients d'ionisation horizontaux linéaires dans un "modèle d'ordre zéro". La figure 1 montre que le "modèle d'ordre zéro", qui représente la distribution verticale locale de la concentration électronique, est le modèle classique de BRADLEY-DUDENEY [11].

Sachant, d'une part que la couche E est très stable et que son ionisation est prévisible avec une grande précision et que, d'autre part elle n'introduit qu'une faible modification des modes de propagation se réfléchissant dans la région F2, nous avons admis que cette couche est constante dans toute la zone étudiée. Les valeurs adoptées pour les paramètres, sont celles qu'ils prennent au centre de la zone (c'est-à-dire à la verticale de la station de sondage), qui se situe à l'origine des coordonnées.

Un autre paramètre a été supposé constant, il s'agit de la demi-épaisseur de la couche F2. Des études ont en effet montré que ce paramètre n'a qu'une influence mineure sur la propagation H.F. [6], aussi a-t-il été fixé à la valeur qu'il prend à l'origine.

Pratiquement les variations géographiques de l'ionisation ont donc été représentées par deux gradients horizontaux linéaires, l'un portant sur la densité maximale d'ionisation de la couche F2, et l'autre sur l'altitude du maximum d'ionisation de cette même couche. Ces gradients, d'amplitudes G_f et G_h sont supposés orientés respectivement dans les directions φ_f et φ_h. Avec ces hypothèses, la concentration électronique dans la couche F2 varie selon la loi :

$$N_2(r,\theta,\varphi) = Nm2(\theta,\varphi) \cdot \left[1 - \frac{r - Rm2(\theta,\varphi)}{Ym2(0,0)} \right] \qquad (1)$$

où :
- $Nm2(\theta,\varphi) = Nm2(0,0) \cdot [1 + Gf.R.\theta \cdot \cos(\varphi - \varphi f)]$
- $Rm2(\theta,\varphi) = (R + hmF2(0,0) \cdot [1 + Gh.R.\theta \cdot \cos(\varphi - \varphi h)]$

avec :
- r,θ,φ coordonnées sphériques du point courant (sondeur au pôle nord du modèle : (R,0,0)
- R rayon de la Terre (6370Km)
- Nm2 concentration électronique maximale (liée à foF2).

b - Modélisation de l'ionogramme -

Le modèle précédemment décrit a permis, grâce à un programme de "tracé des rayons", de simuler l'ionogramme de rétrodiffusion pour l'ensemble des combinaisons des paramètres ionosphériques usuellement rencontrés.

Diverses tentatives d'ajustement de l'ionogramme par des fonctions variées, ont montré que si on trace celui-ci dans un plan aux échelles semi-logarithmiques, la courbe obtenue présente un point d'inflexion, donc une partie quasi-rectiligne, ainsi qu'une courbure très faible [6]. Au niveau de la modélisation, comme au niveau de l'exploitation, cette rectitude apparaît clairement (figure 2), et permet de caractériser l'ionogramme par 2 paramètres. Nous avons posé :

$$\log T_F = A \cdot F + Bo = Ao \cdot X + Bo \qquad (2)$$

avec : $X = F/foF2(0,0)$ fréquence réduite

Ao, Bo pente et ordonnée à l'origine de la tangente au point d'inflexion (ces paramètres sont indépendants de foF2).

c - Recherche d'une méthode d'inversion -

A ce stade de l'étude, le problème de l'inversion peut être formulé mathématiquement de la façon suivante. L'ionosphère est caractérisée par 7 paramètres : foF2, hmF2, ymF2, Gf, Gh, φf et φh. Les ionogrammes sont représentés par l'ensemble des couples (Ao, Bo) associé à l'ensemble des directions de sondage φ. Le problème se ramène donc à la détermination de 7 paramètres à partir d'un nombre au moins aussi élevé d'ionogrammes (couples Ao, Bo).

Remarquons que 3 de ces paramètres : foF2, hmF2 et ymF2 caractérisent le "profil d'ordre zéro", et qu'en conséquence, ils peuvent être mesurés directement grâce à un sondeur zénithal inclus dans la station centrale. Les 4 autres paramètres définissent le "profil d'ordre un", et ne peuvent donc être mesurés que grâce à un sondeur à rétrodiffusion.

Les résultats des sondages peuvent être représentés sous différentes formes, l'une des plus intéressantes consistant à tracer le temps de focalisation en fonction de l'azimut de sondage φ, avec la fréquence en paramètre. Les courbes obtenues, appelées "sondages panoramiques" ou "PPI", ont été simulées pour de nombreuses combinaisons des paramètres Gf, Gh, φf et φh. La figure 3 montre qu'elles possèdent toutes un axe de symétrie et une forme ovale, fermée aux basses fréquences, et ouverte aux hautes fréquences, et celà quelles que soient les directions et les intensités des gradients horizontaux.

Cette dernière remarque est fondamentale, car elle montre qu'il n'est pas possible, à l'aide des seuls ionogrammes et sondages panoramiques, de déterminer les directions et les intensités des 2 gradients horizontaux introduits dans le modèle (*).

(*) Nous avons cependant montré dans une autre étude [10] que, dans l'hypothèse d'une ionosphère de CHAPMAN [12], les gradients horizontaux sont alignés, et qu'il est alors possible de déterminer les intensités et les directions des gradients.

Pratiquement celà signifie que des combinaisons différentes de gradients horizontaux conduisent à des effets observés comparables. L'exemple de la figure 4 illustre cette remarque dans le cas des ionogrammes. C'est la raison pour laquelle nous avons introduit la notion de "gradient unique équivalent", qui représente par définition le gradient d'ionisation (Gfue, φfue), ou d'altitude du maximum d'ionisation (Ghue, φhue) qui agissant seul, produirait les mêmes "effets observés" que les gradients réels.

Si on se limite à la recherche d'un gradient unique équivalent, le nombre de paramètres à mesurer par rétrodiffusion n'est plus que de 2. La direction du gradient (φfue ou φhue) se confond bien évidemment avec l'axe de symétrie des sondages panoramiques, tandis que son intensité (Gfue ou Ghue) est directement liée à la pente de l'ionogramme de rétrodiffusion relevé précisément dans la direction du gradient.

d - Méthode d'inversion proposée -

La méthode finalement retenue découle des remarques précédentes. Elle comporte essentiellement 4 étapes :

- 1ère étape : mesure de foE, foF2, hmF2, ymF2 grâce à un sondeur zénithal, ce qui définit le "profil d'ordre zéro".

- 2ème étape : mesure de la direction (φe) du gradient équivalent, grâce à des sondages panoramiques à plusieurs fréquences. La direction φe est ensuite calculée à partir d'un modèle simple consistant à représenter chaque sondage panoramique par un développement en série de Fourier limité au fondamental :

  $$T_F(\varphi) = a0 + a1.\cos\varphi + b1.\sin\varphi \qquad (3)$$

- 3ème étape : mesure des paramètres (Ao, Bo) de l'ionogramme de rétrodiffusion relevé dans la direction φe (cet ionogramme peut aussi être déduit des sondages panoramiques), puis déduction de Gfue et de Ghue. Pratiquement, une étude antérieure [9] a permis de montrer que les intensités des gradients équivalents ne dépendent, pour un profil d'ordre zéro donné, que de la pente Ao de l'ionogramme. Des formules analytiques donnent alors directement Gfue et Ghue en fonction de Ao.

- 4ème étape : cette étape, non indispensable en théorie, se justifie par l'expérience. Nous savons en effet qu'un gradient d'ionisation, ou de hauteur du maximum d'ionisation, n'apparaît jamais seul, mais en combinaison avec l'autre type de gradient. Nous avons montré d'autre part que des "gradients équivalents", dans le sens où ils conduisent à un ionogramme de mêmes paramètres (Ao, Bo), n'affectent pas la propagation de la même façon à toutes les fréquences. Ce phénomène est rappelé par la figure 5 qui montre que les ionogrammes associés aux gradients uniques n'ont pas la même étendue aux hautes fréquences. Il nous a donc paru judicieux de chercher à réduire l'erreur commise en adoptant simplement Gfue ou Ghue, par l'introduction d'un gradient mixte (Gfm, Ghm) plus conforme aux données physiques.

Plusieurs solutions sont possibles. L'une d'elles consiste à adopter comme relation entre Gfm et Ghm, la relation qui existe dans le modèle de CHAPMAN :

$$\frac{Gfm}{Ghm} = -\frac{R + hmF2}{YmF2} \qquad (4)$$

et à résoudre l'équation Ao = f(Gfm, Ghm) en tenant compte de (4).
Une autre solution peut être envisagée, qui consiste à remplacer (4) par une relation empirique déduite d'une étude statistique des gradients.

Remarquons que si tout l'ionogramme de rétrodiffusion était toujours disponible, l'étape 4 ci-dessus serait inutile. En pratique cependant, la partie haute fréquence de l'ionogramme n'est pas en totalité accessible à cause du rayonnement médiocre des aériens aux faibles angles, ainsi que de la diminution de l'intensité du signal rétrodiffusé aux incidences rasantes, ce qui justifie la méthode proposée.

II.5. Commentaires et perspectives.

La méthode d'inversion proposée a été testée sur un ensemble de cas avec des résultats satisfaisants. Une expérimentation de grande envergure en cours doit permettre d'évaluer les possibilités ultimes de la méthode.

Nous pensons de plus, qu'il est possible d'améliorer la précision, et en particulier que l'on doit pouvoir séparer rigoureusement les deux types de gradients horizontaux grâce à l'introduction d'informations supplémentaires. Ces informations pourraient fort bien provenir d'un sondage en élévation, ainsi que le montre la figure 6.
Il apparaît en effet que des couples (Gf, Gh) équivalents du point de vue de l'ionogramme, ne le sont plus du point de vue du sondage en élévation. La discrimination entre des couples de gradients équivalents est donc en théorie possible, quoique nécessitant des aériens de grandes dimensions.

III. - MODELISATION DU BRUIT ET BROUILLAGE -

Les communications par ondes décamétriques sont caractérisées par leur grande portée. Le test d'un canal, en terme de bilan de propagation et de trajets multiples, n'est qu'un des paramètres de bon établissement d'une liaison. Le brouillage et le bruit au lieu de réception constituent le second paramètre important, puisque le rapport signal/(bruit + brouilleurs) conditionne la probabilité d'erreur dans les liaisons numériques, ou la qualité d'une information transmise sous forme analogique.

Il est cependant établi que le brouillage et le bruit sont présents en permanence et variables dans l'espace et le temps.

III.1. Corrélations spatiales du brouillage.

L'encombrement de la bande décamétrique est typiquement représenté par les relevés de la figure 7, montrant un spectre diurne et un spectre nocturne relevés en Europe occidentale. On y distingue un continuum et des bandes fortement occupées, réservées à la radiodiffusion.

Ces spectres, analysés avec de meilleures résolutions, font apparaître un niveau minimum correspondant au bruit naturel et artificiel entre les spectres, dû aux émissions (figure 8a). La diversité spatiale de l'encombrement spectral est rendue par l'enregistrement de la figure 8b effectué à la même période dans une bande d'analyse de 200KHz en Polynésie française (Tahiti) et en France métropolitaine en Juillet 1988. Ces deux spectres, qui correspondent aux encombrements extrêmes, brouillage très faible en Polynésie, très fort en Europe occidentale, illustrent les différences auxquelles on peut s'attendre.

Ainsi, dans le cas d'une liaison Paris-Papeete, la connaissance de l'encombrement spectral à Paris, lieu d'émission, ne permet pas d'obtenir des informations sur celui de Papeete, lieu de réception.

Une étude effectuée sur des distances plus courtes [13] en 3 lieux d'observation situés aux sommets d'un triangle de côtés 500, 700 et 1000 Km, montre encore une décorrélation importante du brouillage.

Les niveaux de brouillage S en fonction du temps, ont été relevés pour chaque station, avec une résolution de 1KHz sur toute la gamme décamétrique (figure 9) et pendant une période d'observation Tobs. Ils ont été comparés à 15 niveaux de seuil Sc distants de 3dB, autour du niveau - 100dBm correspondant au bruit naturel. On a déterminé alors pour chaque station i :

- Le nombre de fois Nc où $Si < Sc$, et la durée de clarté Tci_n durant laquelle cette condition est respectée.

- La durée moyenne $Dci = \overline{Tci_n} = Tci/Nc$ pendant laquelle, pour la période d'observation : $Si < Sc$, ainsi que la probabilité $Pi(Sc) = Tci/Tobs$ de cet évènement.

- L'écart-type $\sigma_{ci}^2 = \overline{(Tci_n - Dci)^2}$ de la fonction de répartition des durées de clarté Tci_n.

La comparaison des stations i et j entre elles a été faite en déterminant :

- La durée de clarté simultanée $Tcij_n$, qui est la durée pendant laquelle $Si < Sc$ et $Sj < Sc$.

- La durée moyenne $Dcij = \overline{Tcij_n}$ pendant laquelle les niveaux de brouillage dans les stations sont simultanément inférieurs au seuil Sc, ainsi que la probabilité $Pij(Sc)$ de cet évènement.

- L'écart-type $\sigma_{cij}^2 = \overline{(Tcij - Dcij)^2}$ de la fonction de répartition des durées de clarté simultanée pour un seuil Sc.

- Un coefficient d'indépendance entre stations $Iij(Sc)$, qui prend la valeur 1 lorsque les niveaux de brouillage Si et Sj sont des variables indépendantes.

De façon tout à fait similaire, on a déterminé la durée moyenne de clarté simultanée Dcijk pour les 3 stations.

Les mesures effectuées ont montré :

- Que les lois de distribution des niveaux de brouillage suivent une loi log normale (figure 10).

- Que les variations du coefficient d'indépendance font apparaître une quasi indépendance des stations pour des niveaux de brouillage supérieurs à - 100dBm (figure 11).

La décorrélation importante entre les stations de mesure éloignées de 500, 700, 1000Km montre donc qu'il n'est pas possible d'effectuer, à distance, une évaluation du brouillage à partir de mesures locales.

III.2. Corrélation temporelle du brouillage.

Les variations temporelles du brouillage ont été étudiées en Europe occidentale par des relevés effectués sur un même site sur des jours successifs, aux mêmes heures. On a effectué ces analyses sur l'étendue de la gamme décamétrique correspondant aux fréquences passantes, avec des largeurs de bande B = 3KHz et en excluant les bandes de radiodiffusion. Avec un horizon d'observation de 5 Jours, on a mesuré la probabilité pour qu'une fréquence libre un jour le soit dans les jours suivants.

La figure 12 donne les histogrammes des probabilités de clarté des fréquences pour un niveau variable de - 85 à - 60dBm.

Pour les périodes considérées - N = 5 jours - si l'on note Ni le nombre de voies analysées qui ont dépassé le seuil i fois durant cette période, l'histogramme donne la probabilité Pr d'occupation du niveau :

$$Pr_i = \frac{Ni}{P}$$

où P est le nombre total de voies analysées.

Il apparaît que pour ce niveau, situé à environ 10dB au-dessus de l'écart-type du bruit naturel, les fréquences libres, c'est-à-dire celles qui n'ont jamais dépassé le seuil, ou occupées, c'est-à-dire celles qui ont toujours dépassé le seuil, sont prédominantes.

Ces histogrammes sont variables en fonction du seuil et la probabilité de trouver des fréquences libres :

$$Pr_o = \frac{No}{P}$$

est évidemment une fonction croissante du seuil.

Pour un utilisateur, le problème consiste à déterminer si une fréquence libre pendant une période d'observation le sera encore le lendemain.

Pour répondre à cette question on a déterminé la probabilité pour qu'une fréquence libre durant J jours le soit encore le lendemain. Cette probabilité dépend du nombre de jours d'observation J et du seuil Sc . La figure 13 donne les variations de cette probabilité en fonction du seuil et pour un nombre J de jours d'observation.

Ces résultats montrent la corrélation temporelle du brouillage. L'observation du spectre pendant un nombre de jours réduit doit donc permettre d'établir des prévisions du brouillage dans des périodes non conflictuelles. Il semble qu'une observation sur 3 ou 4 jours soit nécessaire pour établir de bonnes prévisions, ce qui permet de penser que les liaisons établies en H.F. se font, pour une part d'entre elles, sous forme de vacations.

IV. - CONCLUSION -

La variabilité du milieu ionosphérique est telle que la gestion des réseaux de communication H.F. impose le recours à des prévisions à très court terme. Celles-ci peuvent être élaborées grâce à une modélisation du brouillage, ainsi qu'à une modélisation de l'ionosphère s'appuyant sur des sondages panoramiques par rétrodiffusion.

La détermination d'un profil d'ionisation moyen, prenant en compte les gradients horizontaux, à partir de l'*ionogramme de rétrodiffusion, est un problème inverse complexe.* Une solution stable a été trouvée, basée sur une double modélisation, celle du profil d'ionisation et celle de l'ionogramme.

Le brouillage présente des propriétés de corrélation temporelle, qui permettent l'évaluation des fréquences libres, avec une probabilité élevée. Une méthode de prévision a été proposée.

L'utilisation conjointe des deux principes énoncés ci-dessus pour la gestion des réseaux de communication H.F., devrait permettre d'en accroître l'efficacité.

B I B L I O G R A P H I E

[1] WRIGHT J.W. - Toward global monitoring of the ionosphere in real time by a bottomside sounding network : the geophysical requirements and the technological opportunity. AGARD Conf. Proc. N° 263 - LISBON - 1979.

[2] PAUL A.K. - *F region tilts and ionogram analysis* - RADIO SCIENCE Vol. 20, N° 4 - July-August 1985.

[3] SABATIER P.C. - Applications de la théorie de l'inversion - Revue du CETHEDEC N° 76 - 1983.

[4] TIKHONOV A., ARSENINE V. - Méthodes de résolution des problèmes mal posés - Editions MIR - MOSCOU - 1976.

[5] RAO N.N. - Inversion of sweep-frequency sky-wave backscatter leading edge for quasiparabolic ionospheric layer parameters - RADIO SCIENCE Vol. 9 N° 10 - 1974.

[6] GOUTELARD C., CARATORI J., COATANHAY J.L., ROLLAND R. - Application de la technique backscatter aux prévisions de transmissions ionosphériques à très court terme - AGARD Conf. Proc. N° 244 - 1977.

[7] HATTFIELD V.E. - *Derivation of ionospheric parameters from backscatter data* - AGARD Conf. Proc. N° 49 - 1969.

[8] DUBROFF R.E., RAO N.N., YEH K.C. - Backscatter inversion in spherically asymetric ionosphere - RADIO SCIENCE Vol. 14 N° 5 - 1979.

[9] GOUTELARD C., CARATORI J. - Modélisation de l'ionosphère dans les problèmes de gestion des réseaux de transmission H.F. - AGARD Conf. Proc. N° 238 - 1978.

[10] CARATORI J., SCHWAB R., GOUTELARD C. - Real time large scale modelling of the ionosphere with backscatter sounding. New results - IEE 4th International Conf. on H.F. radio systems and techniques- LONDON - 1988.

[11] BRADLEY P.A., DUDENEY J.R. - A simple model of the vertical distribution of electron concentration in the ionosphere - J.A.T.P. Vol. 35 pgs 1232-1246 - 1973.

[12] CHAPMAN S. - The absorption and dissociative or ionizing effect of monochromatic radiation in an atmosphere on a rotating earth - Proc. Phys. Soc. LONDON - Vol. 43 Part I : January 1931 - Part II : October 1931.

[13] GOUTELARD C., VAN UFFELEN J.P. - Influence des variations ionosphériques sur les systèmes H.F. à haute fiabilité utilisant de grandes bases - AGARD Conf. Proc. N° 441 - 1988.

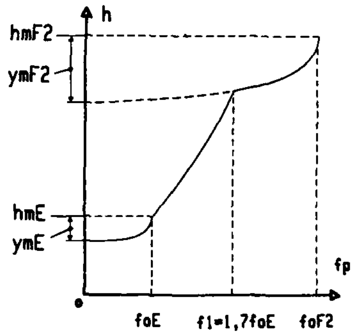


Figure 1 : Profil vertical local
Modèle de BRADLEY-DUDENEY

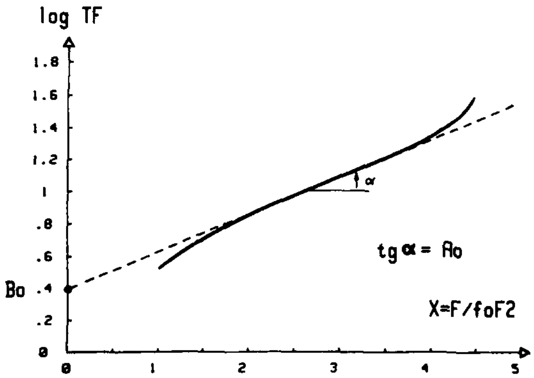


Figure 2 : Modèle d'ionogramme

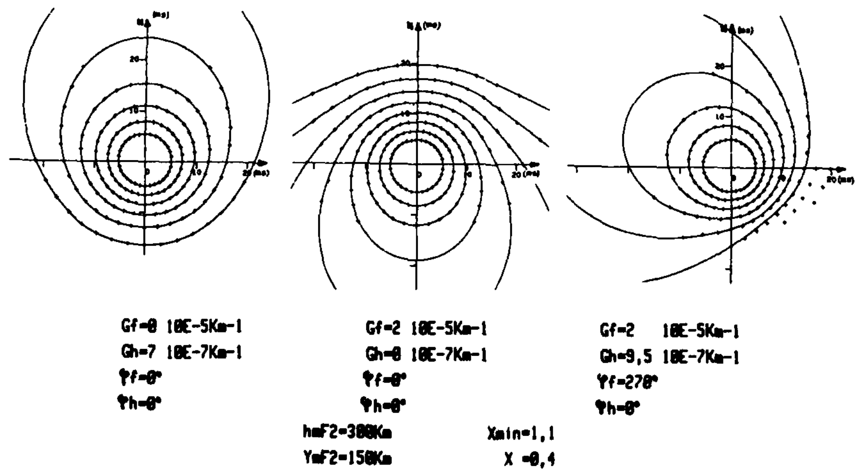
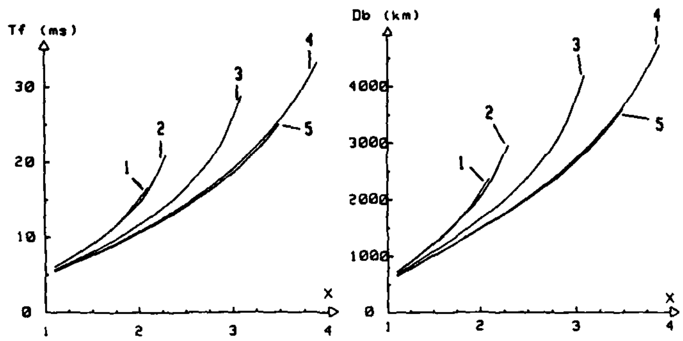


Figure 3 : Modèle de sondage panoramique



hmF2=350 Km	Courbe n°	1 Gf=-2 10E-4Km-1 Gh= 2,6 10E-6Km-1
YmF2=210 Km	Graph n°	2 Gf=-1 10E-4Km-1 Gh= 5,2 10E-6Km-1
		3 Gf= 0 10E-4Km-1 Gh= 0 10E-6Km-1
		4 Gf= 3 10E-4Km-1 Gh= 2,6 10E-6Km-1
		5 Gf= 2 10E-4Km-1 Gh= 0 10E-6Km-1

Figure 4 : Ionogrammes équivalents

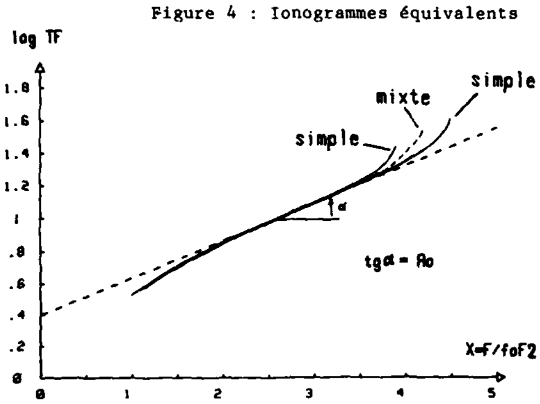


Figure 5 : Réduction de l'erreur grâce au gradient mixte équivalent

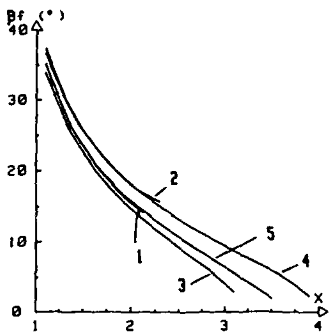
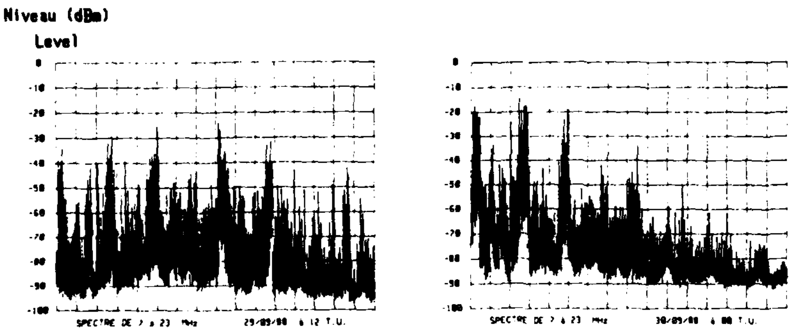


Figure 6 : Levé de l'ambiguité sur les ionogrammes équivalents



a) Jour

b) Nuit

Figure 7 : Encombrement de la bande décamétrique en Europe occidentale

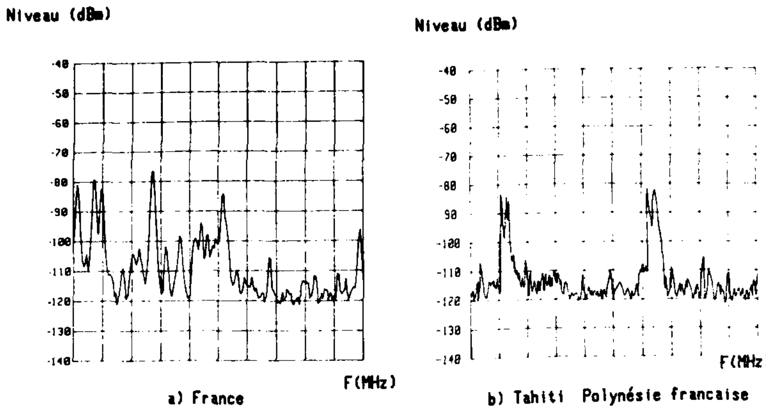


Figure 8 : Comparaison de l'encombrement spectral
22 juillet 1988 - 12h locale - Bande 12,2 à 12,4MHz

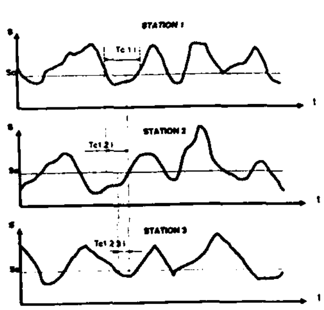


Figure 9 : Analyse du brouillage

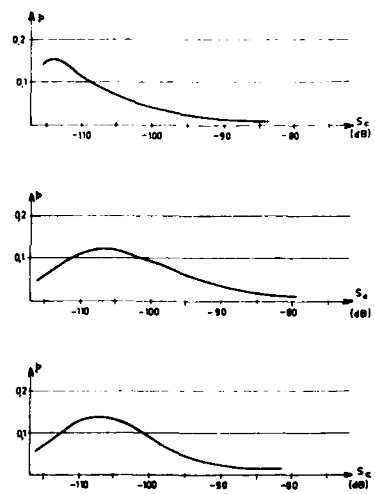


Figure 10 : Densité de probabilité du brouillage

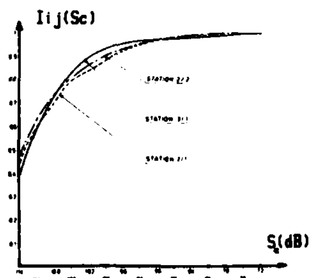


Figure 11 : indépendance entre stations

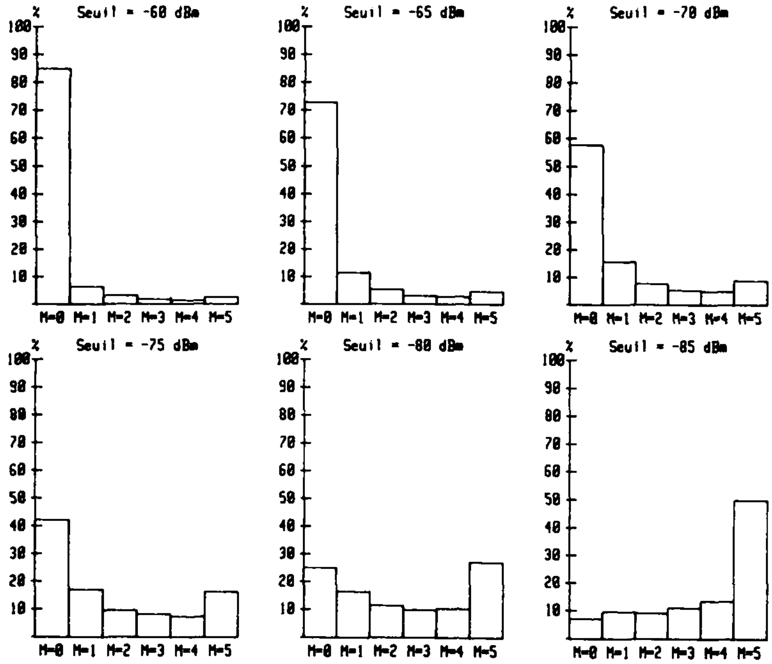


Figure 12 : Histogrammes des probabilités de clarté
sur M jours consécutifs, en fonction du niveau

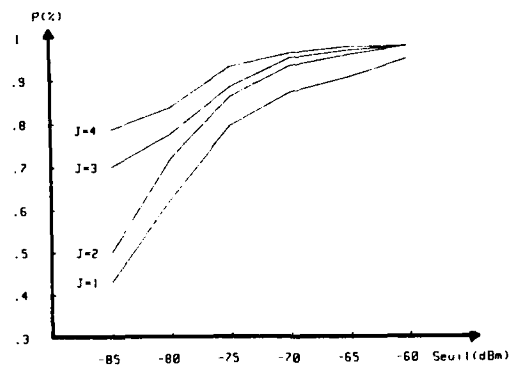


Figure 13 : Probabilité pour qu'une fréquence libre les jours 1 à J
soit encore libre le jour J + 1

DISCUSSION

G.H. HAGN, US

The modeling of HF other-user interference, which you have called jamming (although the interference is not intentional), in a region such as Western Europe is quite important. Such a model is needed to properly predict the performance of modern HF systems, especially adaptive frequency-hopping systems. Empirical models, such as the UMIST model of Profs. P. Laycock and G. Gott, for HF subband congestion at one location could be generalized to cover a region, but more sampling points are needed. Since there is a certain cost per measurement location, the cost to develop such a regional model depends on the dimensions required for the sampling grid. Your data on decorrelation distance suggest that samples should be taken at a distance less than 500 Km in Western Europe. Could you speculate on the separation distance requirement for acquiring the empirical data to develop such a regional model?

AUTHOR'S REPLY

The curve in figure 11 shows that the decorrelation (or independance) between stations depends on the level of comparison adopted.

For a level of 15 dB above the noise, this figure shows a strong decorrelation for distances greater than or equal to 500 Km.

We can speculate that samples at these distances are independant and that shorter distances would be necessary for modelling interferences in Western Europe.

COMMUNICATION OVERVIEW PROGRAM (COP)

D. J. Fang, I. Benoliel
MTL - Washington Division
Massachusetts Technological Laboratory
4300 Montgomery Ave., Suite 204, Bethesda, MD 20814, U.S.A.

and

H. Soicher
U. S. Army Communications-Electronics Command
AMSEL-RD-C3-TR-4, Fort Monmouth, NJ 07703-5202, U.S.A.

ABSTRACT

In establishing a communications system for fulfilling a given mission requirement, multiple levels of decision-making processes are encountered. The decision making process at the lowest level, such as subsystem design and component specification can be accomplished by known routines and procedures, or even by computer packages known as CAD. The middle level of decision-making involves professional articulation as the processes generally involve trade-off's and optimization. As such, engineering prejudice, preference and experience play a significant role. The highest level of decision making, because of its nature, generally is more policy oriented rather than technically oriented. In this level, a broad view including the background, the finance, priority and other administrative concerns are included for decision. The unfortunate part of this is that the decision makers often do not have the necessary insight in making a proper technical decision which has a profound and unequivocal impact on the implementation at lower levels.

The COP Program is being developed for applications at the highest level of decision-making. The program is not a design tool but is intended to provide the necessary scenarios and technical insights to assure a correct decision can be made for the establishment of a communications system to carry out a given mission.

1. INTRODUCTION

The Communication Overview Program (COP) is being developed to provide an electromagnetic assessment tool to be used in planning and implementing C3 and data transmission systems under the Army 21 Concept.

In establishing a communications system for fulfilling a given mission requirement, multiple levels of decision-making processes are encountered. The multi-level hierarchy can be briefly illustrated in Figure 1. The decision making process at the lowest level, such as subsystem design and component specification can be accomplished by known routines and procedures, or even by computer packages known as CAD. The middle level of decision-making involves professional articulation as the processes generally involve trade-off's and optimization. As such, engineering prejudice, preference and experience play a significant role. The highest level of decision making, because of its nature, generally is more policy oriented rather than technically oriented. At this level, a broad view including the background, the finance, priority and other administrative concerns are included for decision. The unfortunate part of this is that the decision makers often do not have the necessary insight in making a proper technical decision which has a profound and unequivocal impact on the implementation at lower levels.

The COP Program is being developed for applications at the highest level of decision making. The program is not a design tool but is intended to provide the necessary scenarios and technical insights to assure a correct decision can be made for the establishment of a communications system to carry out a given mission.

Since typical telecommunication missions can be accomplished by a number of transmission modes such as surface wave, ionospheric wave, LOS, troposcatter, satellite and others, a different system choice would have vastly different performance implications. The essence of the COP Program is to provide to the decision maker a general overview of what a typical choice would provide in terms cost, performance, availability,

expendability, C3 concerns as well as other major concerns related to a mission.

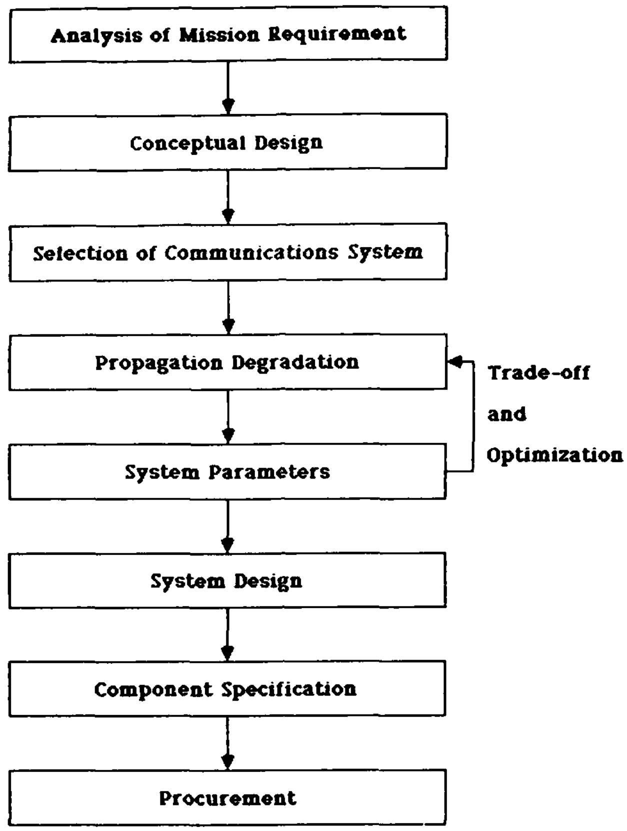


Figure 1. Decision Making Mutilevel Hierarchy

2. COP OBJECTIVES

The design objectives for COP are :

(a) A simulator tool which can readily provide link layout and develop environmental conditions that may induce link performance impairments

(b) An analytical tool for performing link performance analysis in presence of the simulated disturbed propagation

(c) A tool to provide an expeditious way to develop the optimum link design parameter

(d) Easy to operate by mouse, no programming knowledge required

(e) Graphic presentation

(f) Friendly questionnaire, supported by selected options

(g) No indepth knowledge of the propagation mechanisms

(h) Use of computer graphic techniques for input and display

3. CONCEPTUAL APPROACH

3.1 Modular Concept

COP is developed to be accessible to any one involved in telecommunication decision making but not necessarily an expert on propagation and communications system. In addition, COP may be useful for experts desiring to model and exercise link designs and propagation modes. To achieve that objective, a modular concept is used. It includes, at least, the following modules :

* **Basic System Module** - Develops the questionnaire for the user to specify the service requirements and the basic definition of the link of concern.

* **Geographic Locations Module** - Provides the latitude and longitude of the link terminals and pertinent geophysical data associated with each location.

* **Path Simulator Module** - Estimates the path length and simulate the path, describing the significant terrain events and RF interfaces. It provides the means for the user to vary the terrain and interface parameters as needed for the simulation.

* **Propagation Module** - Simulates the propagation events that for a given propagation mode impact the link performance.

* **Modulation and Coding Module** - Provides performance analysis and trade-off of the modulation and coding for better matching the propagation mode capabilities with the required service.

* **Output Module** - Develops the output to produce the information in the most appropriate forms for practical application.

Conceptually, the modulation and coding module is a black box that performs the analysis using the input from the simulators, without *the need to know how they were* generated.

The center piece of the program is the assessment of the degradation of the link performance induced by the simulated environmental conditions. Any radio link performance can be accessed in terms of the *carrier to noise density* (C/N_o). In the COP, appropriate algorithm for the evaluation of C/N_o will be employed. The output module will process it for displaying the relevant performance parameter, which may be S/N (*signal-to-noise in the voice circuit*), the *bit-error-rate* (BER) or the thresholds for the various levels of quality or other parameters of the system. The output is parametric, displayed either as parametric curves giving the time distribution of the performance or as a statement providing the result for a specific condition.

The key for COP are the analytical models and the associated database. The models for the environmental physics and the impairments being implemented are selected from those created by CCIR, NTIS, DCS or other government agencies or by private researchers. The selection criteria are general consensus on model, simplicity of implementing and handling, and accuracy compatible with the basic purposes of COP.

3.2 Propagation Modeling

The electromagnetic propagation is a complex process. To keep the COP as a tool easy to handle, it is necessary to limit the number of variables. A few simplifications are made, which are in line with current practice in the industry. The simplifications concern:

(a) Attenuation

The attenuation induced by a specific path parameter impairment will be specified as the received power degradation induced by that particular impairment as compared with the power received in the absence of the impairment.

(b) Waveform degradation induced by the path

All practical propagation modes have limited bandwidth capability, due to frequency selective mechanisms generating amplitude, group delay and frequency distortions. The resultant signal distortion may become the dominant factor in the link performance degradation, which cannot be improved by increasing the path gain. As any other

characteristics of the electromagnetic propagation, the path bandwidth is time distributed. It depends on the signal modulation and the signal processing at the receiver. However, for given modulation and signal processing, the path bandwidth model can be simplified as a sole function of the multi-path differential time delay and that simplification will be used. Also, for most of the propagation modes, the criterion is only "go" or "no go" and the program will not attempt to estimate the degradation when the information bandwidth exceeds the path bandwidth as defined by the criterion. The path bandwidth is defined in the modulation and coding module and not by the propagation module.

(c) **Signal degradation induced by the equipment**

COP assumes ideal terminal equipment. The degradation induced by real equipment is accounted for by allocating to the equipment a fixed fraction of the total circuit allowable noise power (case of analog circuit) or, in the case of digital links, by adding an implementation margin to the theoretical energy per bit-to-noise density (E_b/N_o) needed for the required BER.

4. MAIN EVENTS OF THE MODULES AND/OR SUB - ROUTINES

4.1 Geographic Location

The location module provides the geographic coordinates for any location "clicked" in a CONUS map or in a world map displayed on the screen. The accuracy is one degree, which correspond to 60 nautical miles or about 111 km. For smaller than 200 km, only one terminal may be "clicked" and a scroll bar is used to enter the path length. It is now being expanded to include for each location, the data on thunderstorm activity, annual mean rain fall and type of climate. A typical presentation is shown in Figure 2.

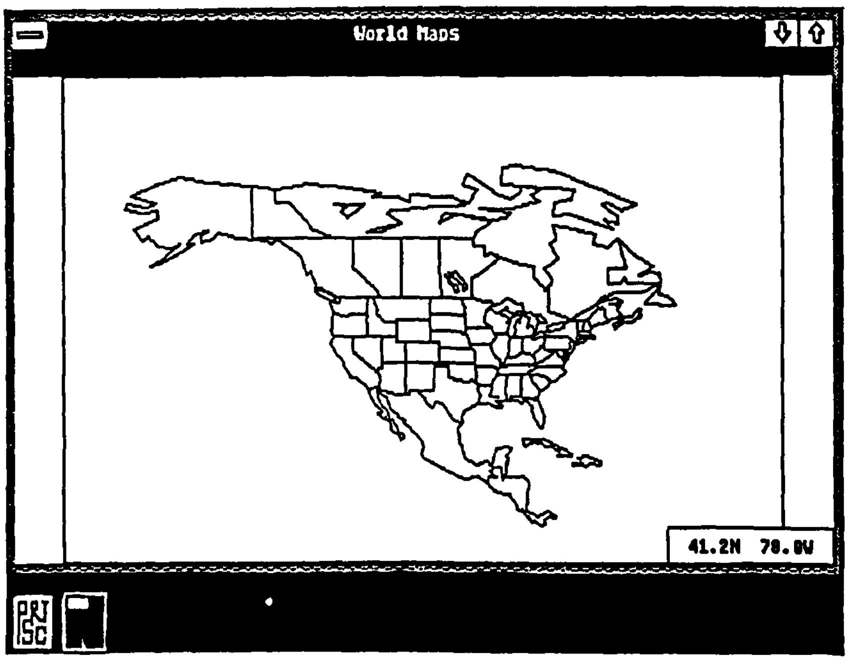


Figure 2. World Map Sub-Routine

4.2 Path Simulator Module

The path simulator uses the output of the geographic location module for estimating the path length and provides for the entry and display of the antenna type and size, antenna height, diversity arrangement, distance, significant obstacles along the path. With the mouse, the user can change any of the path and the RF interface parameters. In

default of path profile data, an expected median terrain will be displayed. This module support the propagation module. A typical presentation is shown in Figure 3.

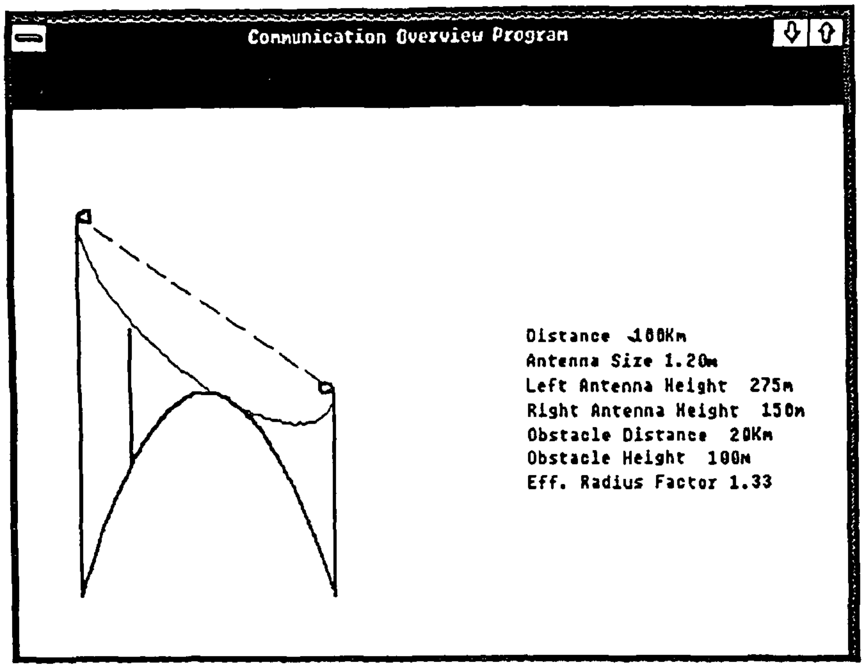


Figure 3. Fresnel Zone and Reflected Path Analysis

4.3 Propagation Module

The propagation module is the heart of COP. It will simulate along the path any of the environmental conditions that may affect the specific propagation mode. The module provides the input to the modulation and coding module.

The disturbing environmental conditions to be simulated are not necessarily the same for all propagation modes. A non-exhaustive list per propagation mode follows:

(a) HF Propagation by Ionospheric Refraction

HF propagation by refraction in the ionosphere is susceptible to any geophysical phenomena that disturb the structure of the ionospheric layers. The effects of the natural phenomena generating fade, multi-path delays, seasonal variations, atmospheric noise, blackouts and other, are well documented and programs providing short and long term predictions are available. One of those programs, the IONCAP, is being incorporated into COP.

(b) Line - of - Sight (LOS)

The degradation effects of the propagation disturbance in *line-of-sight* (LOS) links are different for analog and digital circuits, the latter being more sensitive to multi-path. COP provides the simulation in the conditions needed to analyze both cases. COP will simulate the following path disturbances:

(i) reflection on the ground;
(ii) reduction of the *line-of-sight* clearance from ground obstacles or earth budge;
(iii) multi-path due propagation over different refractive paths;
(iv) out of focus effects due to change in the angle of arrival;
(v) shadow zones created by ground and elevated ducts; and
(vi) absorption by hydrometeors, sand and dust storms.

With the exception of the latter, all other effects are related to the vertical and/or horizontal gradient of the refractivity N, which is a scaled measurement of the refractive index of the air. COP allows the user to simulate a variety of refractive index structures by simulating N profiles and using models developed by NTIS and CCIR, for estimating the disturbed path parameters related to the generated propagation mechanisms. Propagation in ducts and the associated fading mechanisms will be simulated. A climatic classification included in models will be utilized as a guide in the assessment of the susceptibility of the geographic location in the formation of ducting and other fading conditions.

For frequencies above 10 GHz, rain attenuation becomes an important factor in the performance of the link and may limit the path length per hop. COP allows the user to simulate the heavy rain intensity and also use its own model for rain intensity distribution based on thunderstorm intensity index and mean annual rain fall. That model permits simulation of the heavy rain distribution for any geographic locations using input from the geographic location module.

Since the LOS propagation performance depends on the atmosphere structure, which is essentially variable, the LOS long term performance is usually estimated from statistical models, which account for type of the terrain, distance, frequency and diversity arrangement. COP will provide a model for that purpose, which is based in CCIR Report 338-5.

(c) **Diffraction Mode**

The diffraction model estimates the loss for a variety of conditions where surface wave propagation is the dominant mode and also, as an input to LOS sub-routine, the attenuation is induced by reduced Fresnel clearance. The sub-routine allows the user to exercise:

(i) effective radius of the earth;
(ii) frequency;
(iii) path length;
(iv) height and location of ground obstacles;
(v) relative permittivity and effective conductivity.

Two models are incorporated. For frequencies below 20 GHz, CCIR Report 715-2 is used. For frequencies above 20 GHz, COP uses a combination of that model with the model proposed in ESSA Technical Report ERL 79-ITS 67 for propagation over irregular terrain. The applicable model depends on the geometry of the obstacle intercepting the propagation path.

(d) **Troposcattering Mode**

The elements that a troposcatter link designer may exercise for better performance are:

(i) elevation angles;
(ii) path length;
(iii) scatter volume (defined by the antenna patterns);
(iv) diversity arrangement;
(v) frequency; and
(vi) signal modulation and processing.

The troposcatter sub-routine does not address the modulation, which is to be analyzed by the Modulation and Coding Module. Some of the applications of the simulator are analyses of the effectiveness of space, frequency, polarization and angle diversity; also, the implication of the antenna location as it relates the elevation angle. The troposcattering module will use Method II of CCIR Report 238-5. This method is easier to implement and is believed to be able to provide results consistent with those obtained from the more sophisticated Method I.

(e) **Satellite Link Mode**

The satellite link sub-routine will simulate:

(i) free space link;
(ii) elevation angle;
(iii) gaseous absorption;

(iv) rain attenuation and depolarization;
(v) gravity waves and troposcattering scintillation effect;
(vi) ground reflection;
(vii) antenna pattern and gain; and
(viii) interference between adjacent satellites.

The gaseous and rain models will be adequate to estimate the atmospheric gaseous attenuation as a function of the frequency and thus provide a tool to assess the performance of frequencies above 20 GHz for space links and to identify the usable frequency bands. The simulation of the ground reflection is of particular interest for mobile systems using small non-directive antennas and also for the case of large antennas at very low elevation angles. For interference analysis, COP will simulate the antenna patterns and orbital separation; it provides C/I and a general assessment of the induced impairment.

4.4 Modulation and Coding Module

In the COP concept, the modulation and coding module performs as a black box, using the propagation parameters generated by the propagation module, without considering the way they were generated. The module will perform the analysis to define the waveform and signal coding for optimum matching of the propagation path or, for given propagation, determine the circuit performance. The module will use some analytical model, but will be fundamentally based on databases derived from parametric curves.

4.5 Output Module

The output module will display the results derived by the Modulation and Coding module in the way selected by user from a menu of options. The results provided by the modulation and coding module are expressed in C/N_o and the output module will process it for display in the desired format. The available formats are C/N_o, carrier to noise density, S/N, signal to noise in the voice circuit, and BER. Those results may be for a given operating point or, if available, by parametric time distribution curves. The input used by the simulators and by the modulation and coding module will be listed.

5. Conclusion

While the COP Program is still under development, its utility for decision-making becomes obvious. In designing the program, we have invited communications systems users who do not have propagation and system background to witness and to operate the COP Program. Based on their reactions and input, we are able to further streamline the input-output procedure, resulting in more involvement of the COP user. Our final objective is to enable the user to carry on intelligent exchanges with the system designer regarding the system and mission requirements, which is the most important element in high level decision-making.

6. Acknowledgement

The two leading authors (*D. J. Fang* and *I. Benoliel*) gratefully acknowledge the significant contributions by their colleagues, Messrs. *C. L. Lo, S. C. Chang* and *S. J. Lin* in making the COP Program possible.

DISCUSSION

G. HAGN, US

Frequently it is important to know what <u>not</u> to do. Your start on developing a very general type of decision aid for radio system selection to meet a set of requirements seems to try to identify one system. It may be more valuable for selecting a limited number of viable system options for more detailed analysis than for selecting the most appropriate system. A rule-based applied artificial intelligence (AI) system could help the planner by pointing out incompatible combinations and denying their selection while determining viable system options.

AUTHOR'S REPLY

No reply needed.

T. K. FITZSIMONS,

In determining the design characteristics of a new system, one of the most constraining factors is the electromagnetic environment. This needs to be determined for both peacetime and wartime operation and may need to include many complex scenarios. For example, a system like SCRA could operate anywhere in the VHF band or in the UHF band and the determination may depend entirely on the existing electromagnetic environment. Would you comment on this please?

AUTHOR'S REPLY

I agree. For the time being, our program only incorporates simple scenarios for general decision purpose. Complex peacetime and wartime scenarios can be incorporated afterwards when we establish the merit of the program.

The Thermal Behavior of Natural Backgrounds and its Prediction by Means of Numerical Models

Fritz G. Wollenweber

German Military Geophysical Office

5580 Traben-Trarbach
Mont Royal
Federal Republic of Germany

ABSTRACT

The Federal Office for Military Technology and Procurement together with the German Military Geophysical Office (GMGO) are conducting a research effort to establish procedures to describe and to predict thermal behavior of natural backgrounds under varying meteorological conditions.

The experimental phase of this effort started in 1988 with measurements of surface temperatures of different backgrounds like trees, crops, pasture, roads and buildings together with meteorological measurements. These measurements are conducted in two geographical locations, the Meppen proving ground in Northern Germany and Oberjettenberg proving ground in the Alps. The measurement campaign has a time frame of at least one year with measurements taking place during three days each month. It is anticipated to gather a data set that reflects the full range of thermal backgrounds for tactically important locations in Central Europe.

One objective of this effort is to compare and evaluate thermal background models that are under development as a companion task. The results of the Meppen '85 experiment, a pilot study to this campaign, are presented. They are compared with predictions of the current version of the US Air Force Tactical Decision Aid (USAF TDA) and with the thermal background model "CANOPY" developed by GMGO. The GMGO model determines energy exchange within canopies with a multi layer approach for energy, moisture and momentum fluxes. In comparison with measurements the GMGO model performs much better than the USAF IR-TDA model. Differences between model results can be explained by the neglect of latent heat flux in the USAF TDA model for vegetation backgrounds and its semi-statistical approach.

1. INTRODUCTION

At the moment the German Army is in the process of changing its structure in what is called the "Heeresstruktur 2000". The Heeresstruktur 2000 heavily depends on air transport capabilities for timely deployment of men and material and on the ability of anti tank helicopter units to engage enemy tank forces. To maintain these capabilities it is necessary to fly helicopters in low altitudes and close to obstacles during day and night under various weather conditions.

The upcoming generation of army helicopters will be equipped with image intensifiers and thermal imagers to cope with the night vision problem. Unfortunately both systems are affected by weather impacts. The German Military Geophysical Office (GMGO) is in charge to establish methods to advice the Army Aviation Corps how and when it will be possible to use electro-optical devices for navigation and for target acquisition. For both purposes information about thermal backgrounds is essential. The task is to establish methods to predict thermal behaviour of different background types, to incorporate atmospheric attenuation and to include system design parameters to obtain overall performance of electro-optical systems.

The knowledge about IR-backgrounds results from measurements of background variability and from theoretical considerations. Measurements are made with radiometers and thermal imagers that are hard to compare in terms of thermal and geometrical resolution. The data are gathered at different geographical and topographical locations and under different climatic conditions. Meteorological data sets that go with these measurements are limited and in some cases are lacking important parameters. But nevertheless it is worthwhile to compile these measurements since this material provides a knowledge base about important features of natural IR-backgrounds.

Theoretical models started by equating energy inputs and outputs of single leaves (Raschke, 1956; Gates, 1969; Linacre, 1967). Energy inputs to the leaf are the sun's irradiance and thermal radiation from sky and from nearby objects. Heat is lost by three processes: re-radiation, convection and transpiration. The resulting equation for a leaf at thermal equilibrium with its surroundings can be solved to yield leaf temperature.

Large differences in thermal behaviour are found between natural and artificial backgrounds. They can be explained by differences in energy partition into different energy fluxes (ground heat flux, flux of latent and sensible heat, and heat storage), reflecting differences in heat capacity and conductivity as well as soil water conditions.

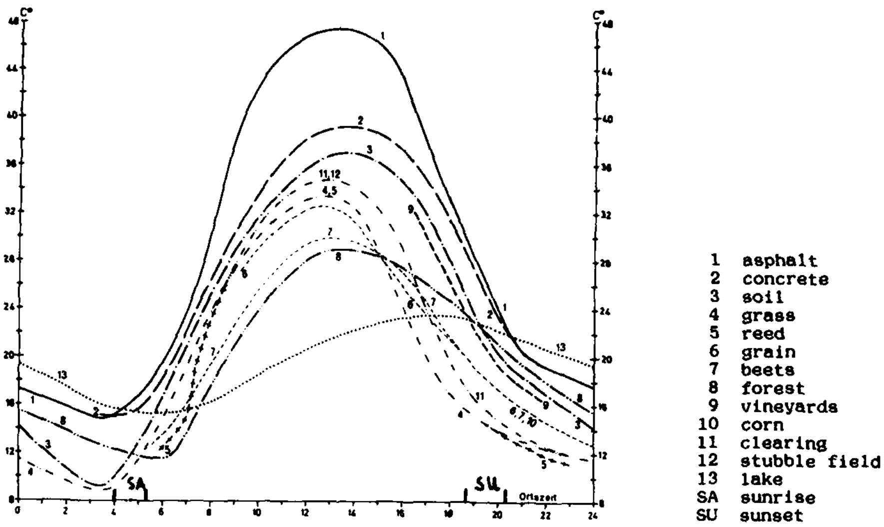


Figure 1: Surface temperature of different background types under radiative conditions for a midlatitude summer day (Fezer, 1975).

Figure 1 shows the daily course of surface temperature of different backgrounds under radiative conditions for a midlatitude summer situation (Fezer, 1975). Around sunrise (SA) and sunset (SU), temperature differences between backgrounds are reduced. Actually in some cases crossover effects can be found. There is a relation between vegetation height and surface temperature. Canopy temperature is modified by vegetation height (Figur 2; Green, Harding and Oliver, 1984). The higher the canopy, the weaker the daily variation of surface temperature and visa versa.

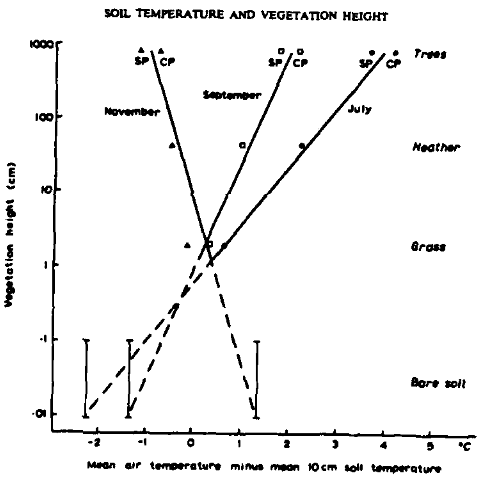


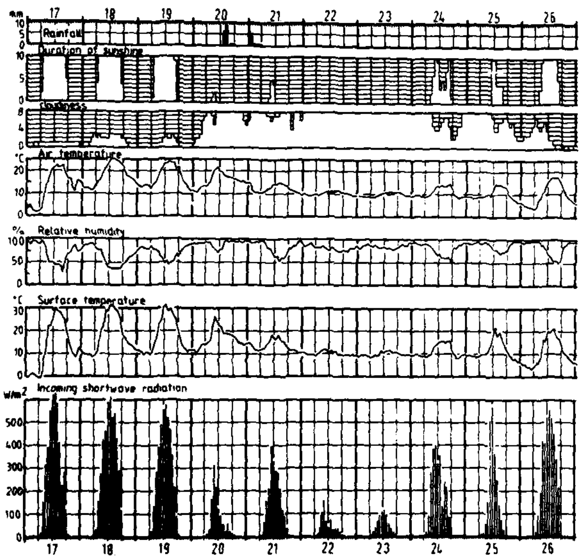
Figure 2: Relation between mean air temperature and mean 10 cm soil temperature for July, September and November 1981 at Thetford (UK); CP = corsican pine, SP = scots pine (Green, Harding, Oliver, 1984).

Vegetation is largely affected by soil water conditions and plant disease. The temperature difference between morning and noon temperatures is used to determine water stress conditions of crops from space (Idso et al.,1976). Table 1 states measured surface temperatures of sugar beet under irrigation and under water stress conditions. For irrigated plants there is only a slight temperature difference between surface and air temperature, whereas the plants under stress conditions show surface temperatures that deviate by several degrees from air temperature.

unirrigated sugar beet				
date	11 August 1976		16 August 1976	
time	13.15		10.30	
air temperature	24.0		20.5	°C
leaf temperature	30.1		28.3	°C
temperature difference	6.1		7.8	°C
irrigated sugar beet				
date	11 August 1976		16 August 1976	
time	13.45	14.45	11.25	
air temperature	24.5	24.6	22.0	°C
leaf temperature	24.8	25.3	23.9	°C
temperature difference	0.3	0.7	1.9	°C

Table 1: Air and leaf temperature of sugar beet with and without water stress (Hoyningen-Huene and Bramm, 1981)

Time series of several parameters measured over a stubble field show the impact of weather conditions on surface temperature (Figure 3). During the first three days radiative conditions prevail with little cloudiness. Accordingly daily variations of air and surface temperatures are well pronounced. When the sky is overcast or it is even raining, both, air and surface temperature show only small daily variations and are almost identical.



Sept. 1979

Figure 3: Daily variation of several meteorological parameters over a stubble field for different weather situations (Jaeger, 1981).

2. MODELING EFFORTS

In this paragraph different approaches to the forecast problem that has to be solved will be outlined. In principle there are three different methods. There is a statistical approach, a first principles model and mixtures of both methods. Lateron the USAF IR-TDA and the GMGO CANOPY model will be compared with measurements.

The model of Mansi and Walls (1987) is an example of the statistical approach. They express the surface temperatures T_s as a linear combination of several important meteorological parameters P_i (Equation 1).

$$T_s = a_0 + \sum_{i=1}^{n-1} a_i P_i \qquad (1)$$

The parameters P_i and coefficients a_i are determined by performing multiple linear regression on about 200 days of observed thermal images (about 40000). This is a very tedious and time consuming approach which is also limited by the fact that the coefficients can not be applied to arbitrary tree or grass areas, since surface temperatures depend on the degree of exposure to the particular element in the scene.

2.1 USAF-TDA BACKGROUND MODEL

The USAF-TDA is an operational tactical decision aid, that is described in detail in several technical reports (for details see Higgins et al., 1987). In this paper we will discuss and evaluate the IR-background/target model of the USAF-TDA.

The USAF-TDA model is based on the equilibrium solution of the heat balance equation (Equation 2). It expresses balance between the following heat fluxes at the surface of the background/target.

1) irradiance on the surface from sun (W_{sun}) and sky (W_{sky})
2) thermal radiation from the surface, σT_s^4
3) free and forced convective exchange of heat, depending on wind speed (V), surface temperature (T_s) and air temperature (T_a)
4) exchange of latent heat (only for certain background types), as a function of vapor pressure (e) and saturated vapor pressure ($e_s(T_s)$)
5) conductive heat exchange between the slab's surface and its interior, modeled on surface temperature (T_s) and core temperature (T_c)
6) the rate of heat storage within the slab, modeled on the rate of ambient temperature change (dT_3)
7) generation of heat by an active target (OSP)

$$\begin{aligned} F = {} & \alpha_1 W_{sun} + \alpha_2(W_{sky} - \sigma T_s^4) + (\alpha_3 + \alpha_4 V)(T_a - T_s) \\ & + \alpha_8/\alpha_3(\alpha_3 + \alpha_4 V)(e - e_s(T_s)) + \alpha_5 F_1(T_c - T_s)/N \\ & - N\,F_2\,dT_3/3 + OSP = 0 \end{aligned} \qquad (2)$$

The USAF-TDA model is a mixture of a statistical approach and a first principles model. Its coefficients α_i, F_i and N, that characterize each target and background were evaluated by the modelers by fitting them to relevant meteorological measurements and to experimental data, typically a 24-hour sequence of radiometric measurements on the particular target or background.

A serious disadvantage of this model is the neglect of latent heat flux for all types of vegetation backgrounds. The flux of latent heat is a major sink or source of energy for vegetation canopies and provision should be made to account for it.

Sun's irradiance W_{sun} is calculated with the Shapiro model (Shapiro, 1982) which is a two stream, plane parallel, broadband model. It is used with three atmospheric layers and a ground surface. Besides date, time, latitude and longitude, cloud type, cloud amount and cloud height are needed as an input.

The Wachtmann model (Hodges et al., 1983) is used to determine sky irradiance (W_{sky}). Here again cloud amount and cloud height for three atmospheric layers are needed together with surface air temperature and water vapor pressure.

2.2 GMGO-BACKGROUND MODEL "CANOPY"

As already mentioned evaporation from vegetation is a major energy flow in vegetation canopies. The GMGO model "CANOPY" (Wollenweber, 1986) accounts for latent heat flux as a function of atmospheric saturation deficit (which is directly related to relative humidity) and of stomatal resistance. To keep computational efforts as small as possible, certain assumptions have to be made to simplify the problem. The main simplification is the assumption that heat capacity is of minor importance for vegetation canopies and consequently can be omitted which eliminates time dependence.

The canopy is divided into 6 layers (Figure 4). For each layer it is assumed, that the heat balance equation reduces to Equation 3, that means the sum of net radiation (Q), sensible heat flux (L) and latent heat flux (V) equals zero.

$$Q + L + V = 0 \qquad (3)$$

The lowest canopy layer with the soil surface as lower boundary has to account for soil heat flux B.

Heat fluxes at the top of each individual layer, marked by index "i", are calculated from Equations 4a–4b.

$$H_i = -\rho c_p (T_{a,i-1}) - T_{a,i})/R_i \qquad (4a)$$

$$V_i = -\rho c_p (e_{a,i-1}) - e_{a,i})/R_i \qquad (4b)$$

$$R_i = \int_z^{z+dz} (1/K(z))dz \qquad z = \text{vertical coordinate}$$

The changes of heat flux within a single canopy layer are modeled according to Equation 4c and 4d.

$$dH_i = \rho c_p (T_{si} - T_{ai})/RH_i \qquad (4c)$$

$$dV_i = \rho c_p (e(T_{si}) - e_{ai})/RV_i \qquad (4d)$$

$$RV_i = RH_i + RS_i$$

T_{ai} = air temperature
T_{si} = leaf surface temperature
e_{ai} = vapor pressure
$E(T_{si})$ = saturation vapor pressure
c_p = heat capacity
ρ = air density

These equations form a set of two coupled linear equation systems, that can be visualized by a resistance network (Figure 4). They can be solved for surface temperatures, air temperatures and humidities (Waggoner and Reifsnyder, 1968; Chen, 1984).

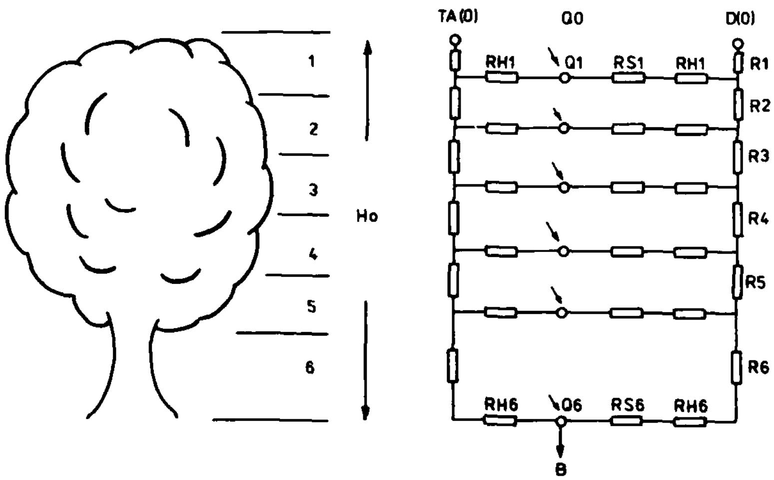


Figure 4: Layer structure of the GMGO CANOPY model and the symbolic representation of the different energy fluxes by a resistance network (symbols are stated in text).

To solve the system further assumptions have to be introduced. Wind velocity and turbulent diffusivity are modeled to increase exponentially with height within the canopy (Equation 5a, 5b).

$$U(z) = UH \exp(-a(1-z/H)) \qquad (5a)$$

H = canopy height

$$K(z) = KH \exp(-a(1-z/H)) \qquad (5b)$$

The coefficient 'a' is varying between 1 and 5 according to canopy type, but the model results do not depend strong on 'a'. UH and KH are wind velocity and diffusivity at the upper boundary (z=H) of the canopy. They are determined from boundary layer theory (Haugen, 1973).

Vertical variation of net radiation in a vegetation canopy is well described by Equation 6 (Monsi and Saeki, 1953).

$$Q(z) = Q_o \exp(-a_s LAI(z)) \qquad (6)$$

LAI(z) = leaf area above z
Q_o = net radiation at canopy top

The extinction coefficient 'a_s' varies between 0.2 and 1.0 (Monteith, 1975) depending mainly on vegetation type and time of day (solar height). 'a_s' can be determined from complex canopy radiation models, e.g. Goudriaan (1977) or Braden (1982).

The soil heat flux is modeled following a suggestion of Nickerson and Smiley (1975) that connects soil heat flux to net radiation (Equation 7). Deardorff (1978) proved the method to be accurate enough compared to more elaborate models.

$$B = \text{const. } Q_6 \qquad Q_6 = \text{net radiation below canopy} \qquad (7)$$

Sensitivity studies showed, that for vegetation canopies soil heat flux is of minor importance as long as most of the soil is covered by vegetation.

Evaporation from soil below the canopy has to be estimated by specifying a Bowen ratio. Fortunately the influence of soil evaporation decreases as leaf area index increases. Only in case of sparse vegetation the Bowen ratio has to be known accurately.

Stomatal resistance depends on several parameters. The most important are water supply, radiation and leaf temperature (Figure 5). The relationship between stomatal resistance RS and radiative flux I (Shawcroft et. al., 1974) and water supply (nFK = plant available soil water content) are given by Equation 8.

$$RS(I, nFK) = RS0(nFK) + C1(nFK)/(C2 + I)$$

$$RS0(nFK) = RS0_{min} - (RS0_{min} - RS0_{max})(100.-nFK)/60. \qquad (8)$$

$$C1(nFK) = C1_o + C1_o(100.-nFK)/60.$$

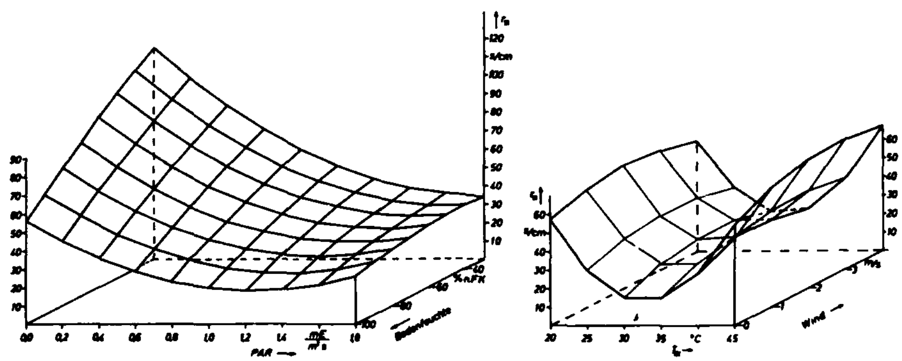


Figure 5: Stomatal resistance RS of corn leaves (v.Hoyningen-Huene and Bramm, 1981).
Left: RS as a function of radiation (PAR) and soil moisture (nFK).
Right: RS as a function of leaf temperature (T_{Bl}) and wind speed.

Finally each individual leaf is surrounded by a boundary layer that poses a certain resistance to transport processes. Theoretical and experimental considerations result in Equation 9 that relates boundary layer resistance with leaf size (L_i) and wind velocity (U_i) (Pearman et. al., 1972).

$$RH_i = 90.(L_i/U_i)^{0.5} \qquad (9)$$

Model results that demonstrate the influence of some important parameters on canopy radiation temperature (ST) and latent heat flux (V) are shown in Figures 6-9. Figure 6 and 7 show a linear relationship between net radiation (QO), relative humidity (U) and canopy radiation temperature (ST), revealing net radiation and relative humidity as forcing functions of energy fluxes. The logarithm of vegetation height H is linearily related to the temperature difference between canopy surface temperature and air temperature (ST-TAO) (Figure 8), which is in agreement with experimental results. Figure 9 shows how canopy surface temperature depends on wind speed. ST shows a very strong dependence on wind speed when wind velocity is low and stomatal resistance is high. This situation is often found during nighttime. As a consequence it is of great practical importance to predict local differences in wind speed since that largely affects thermal background behavior.

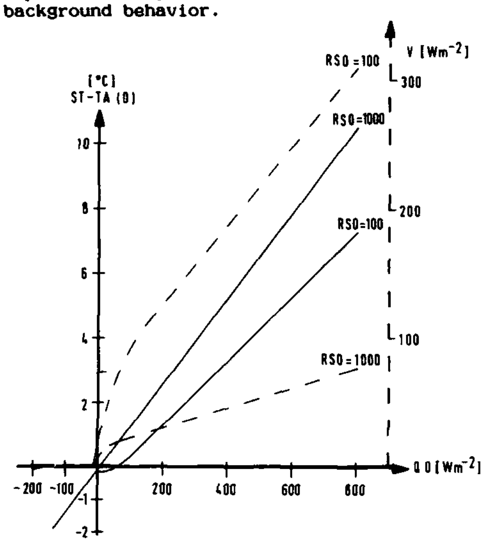


Figure 6: Dependence of latent heat flux, V and temperature difference between canopy temperature ST and air temperature TA(0) on net radiation QO, for two stomatal resistances RSO.

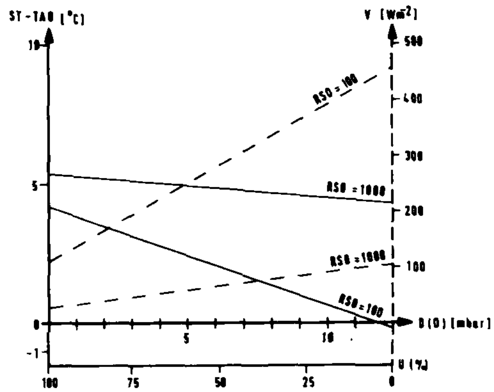


Figure 7: ST-TA(0) and V versus relative humidity U and saturation deficit D(0) for two stomatal resistances RSO.

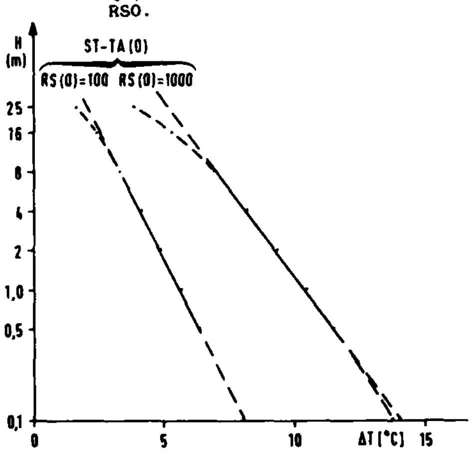


Figure 8: Dependence of radiation temperature (ST) on canopy height (H) (the difference ST-TA(0) is shown) for two stomatal resistances RS(0)

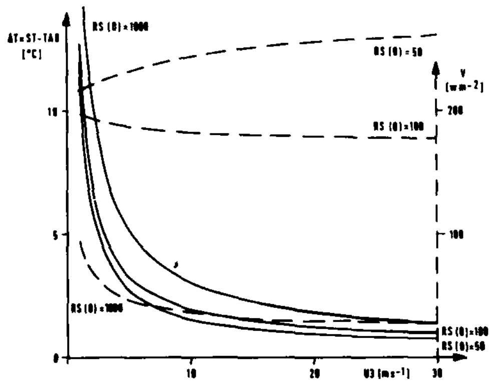


Figure 9: ST-TA(0) and V as a function of wind speed U3 (wind speed in 30m above ground) for three values of RS(0).

3. IR-BACKGROUND MEASUREMENTS

In 1985 GMGO conducted a pilot study of IR-background measurements at Meppen proving ground in Northern Germany. The experiment was scheduled for two weeks in May 1985. The proving ground is equipped with various meteorological instruments and standard measurements are available. In addition to this a limited set of local observations was made at the measurement sites. Several background types were defined and the thermal imaging system (AGA Thermovision 780) was installed on a truck to allow transportability from one site to another. To move the system took about half an hour and consequently measurements were not made at the same time. In that way about six measurements (thermal images) a day were gathered for each background type. The measurements are unevenly distributed over the day with a poor time resolution. Because of weather conditions in May '85 only two days of data are available.

The Federal Office for Military Technology and Procurement and GMGO used the experiences of this pilot study and set up an extensive IR-measurement program, that started in 1988 in Meppen and Oberjettenberg. The positions of the measurement sites are marked in Figure 10.

The measurements take place on three days each month. The sites are chosen so that several backgrounds can be studied from one position of the AGA system. In the vicinity of the measurement sites weather stations are deployed to automatically record the important meteorological parameters. Every two hours thermal images of all backgrounds are taken. One hour before and after sunset and sunrise the timestep is reduced to half an hour for better resolution of the transition from nighttime to daytime conditions. The time frame of the overall experiment is one year.

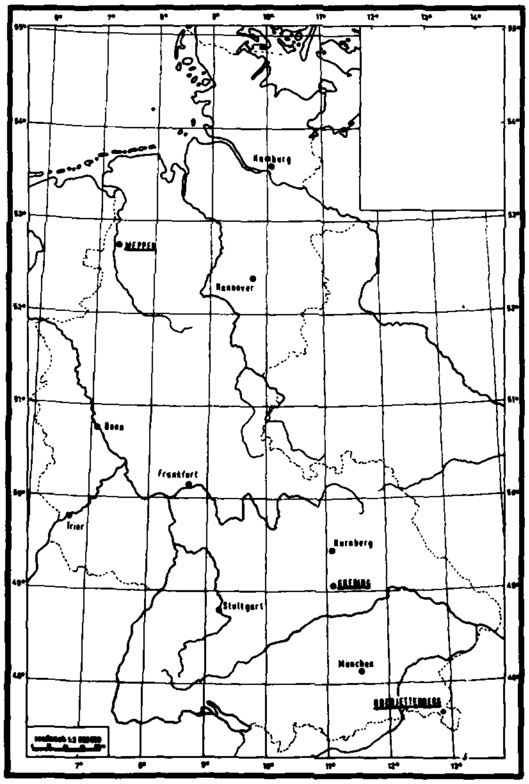


Figure 10: Location of the measurement sites Meppen (Northern Germany) and Oberjettenberg (Alps).

4. COMPARISON OF USAF-TDA AND GMGO CANOPY-MODEL WITH MEASUREMENTS

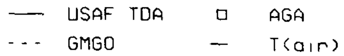
To evaluate the IR-background models of the USAF-TDA and the GMGO CANOPY model, the results of these models and the AGA measurements of canopy surface temperature of the Meppen '85 experiment are compared. Measured values of air temperature, relative humidity, wind speed, visibility, cloud amount, cloud type and cloud height are model inputs.

In Figures 11-13 daily variations of radiation temperature of different canopy types are shown for 29 May 1988, as predicted by the USAF-TDA and the GMGO CANOPY model. These results are compared with the AGA-measurements. For comparison the measured air temperature is also shown. It can be seen that the GMGO model performs much better than the USAF-TDA model does. The reason for this behaviour is, that the parameters of the USAF-TDA model are matched to a specific scene. Apperently the USAF-TDA is not flexible enough to adjust to other situations (information about USAF-TDA backgrounds and thermal images used to adjust model parameters are given in AFGL, 1988). In agreement with these results, Leidner et al. (1988) found a poor performance of the unadjusted USAF TDA models compared to measured surface temperatures.

Figures 14-15 show differences in radiation temperature between different canopy types. Although the USAF-TDA model performs better here than it did in terms of absolute values, the results are still not satisfying.

Figures 16-18 show the same sequence of diagrams, except for other vegetation types and for 30 May 1985.

Figures 11-13: Measured canopy temperatures (AGA), air temperature (T(air)) and model predictions of USAF IR-TDA model and GMGO CANOPY model versus time; Meppen, 29 May 1985.



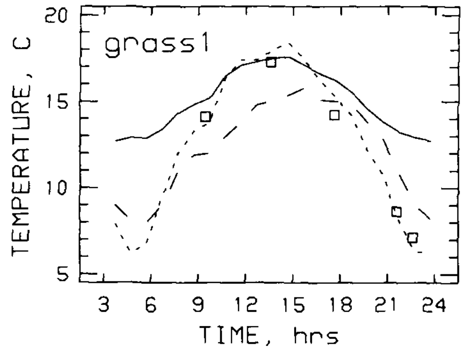


Figure 11: GRASS1 - mowed grass, Meppen, 29 May 1985.

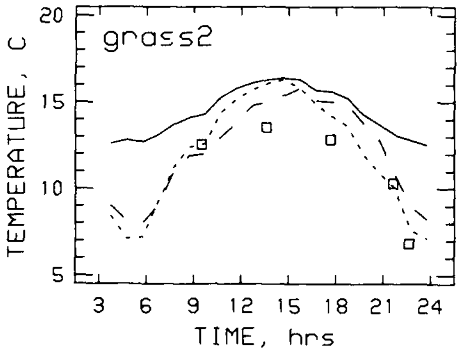


Figure 12: GRASS2 - tall grass, Meppen, 29 May 1985.

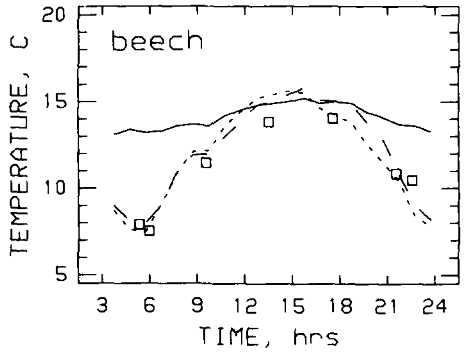
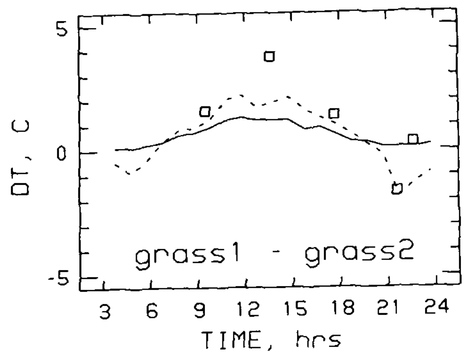
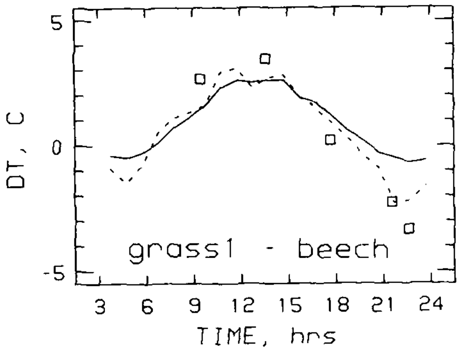


Figure 13: BEECH, Meppen, 29 May 1985.





—— USAF TDA   □ AGA
- - - GMGO   — T(air)

Figure 14: **Temperature difference GRASS1 -GRASS2, Meppen, 29 May 1985.**

Figure 15: **Temperature difference GRASS1 -BEECH, Meppen, 29 May 1985.**

Figures 16-18: **Measured canopy temperatures (AGA), air temperature (T(air)) and model predictions of USAF IR-TDA model and GMGO CANOPY model versus time; Meppen, 30 May 1985.**

—— USAF TDA   □ AGA
- - - GMGO   — T(air)

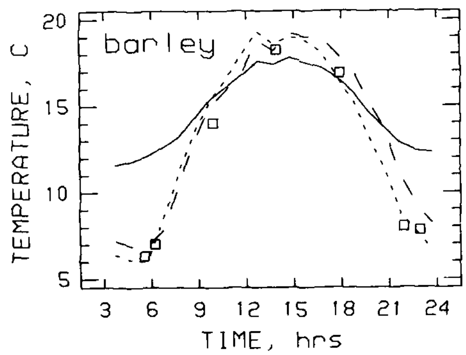


Figure 16: **BARLEY, Meppen, 30 May 1985.**

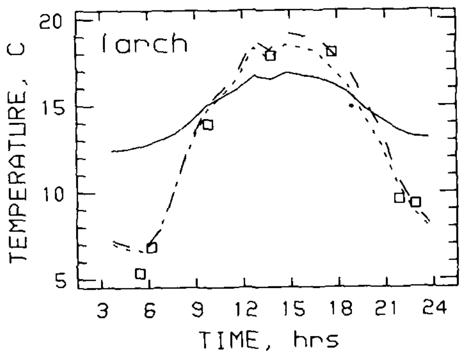


Figure 17: **LARCH, Meppen, 30 May 1985.**

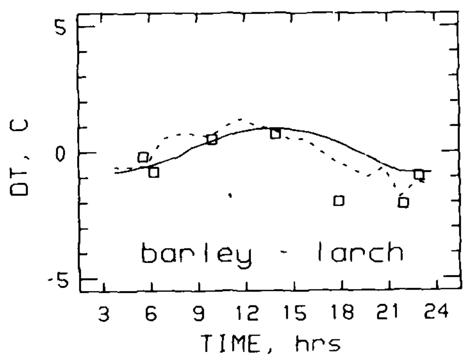


Figure 18: **Temperature difference BARLEY -LARCH, Meppen, 30 May 1985.**

5. FURTHER DEVELOPMENT

GMGO's intention is to develop a decision aid to predict FLIR performance. Atmospheric attenuation will be addressed by the transmission model LOWTRAN6 and sensor characteristics will be accounted for by use of the appropriate MRTD function. There will be three major milestones. The first milestone will be the development of a semi-manual version. This version will allow data and forecasts to be input manually. Terrain is still considered to be flat. Milestone #2 will include wind modification by topography and also variation of solar heating by terrain inclination. Data and forecasts still have to be input manually. In milestone #3 the decision aid will be linked to GMGO's mainframe computer to get data and predictions out of databases and out of numerical models. Model predictions will be taken from GMGO's mesoscale model, that has a horizontal grid size of 63.5 km. The local weather forecaster will have an option to manipulate data to correct errors in data or wrong forecasts. The corrected data and model predictions will be used as an input to an analytical small scale model that interpolates the mesoscale predictions to a smaller scale (grid size 1 km) and also accounts for topographical influence on wind. Finally a one dimensional boundary layer model will be used to predict local values of temperature, humidity and wind speed. While the TDA versions of milestones #1 and #2 are PC/AT based TDA's the final version will need a workstation type computer (e.g. HP 9000 or VAX).

6. REFERENCES

AFGL, "Mark II Tactical Decision Aids for Mikrocomputersystems - Users manual (Scientific Rep. No 34), 1988, AFGL, Hanscom AFB, MA

Braden, H., "Simulationsmodell für den Wasser-, Energie- und Stoffhaushalt in Pflanzenbeständen", 1982, Dissertation Universität Hannover

Chen, J., "Uncoupled multi-layer model for the transfer of sensible and latent heat flux densities from vegetation", Bound. Layer Met., 28, pp. 213-225, 1984

Deardorff, J. W., "Efficient prediction of ground surface temperature with inclusion of a layer of vegetation", J. Geophys. Research, 83, pp. 1889-1904, 1978

Fezer, F., "Lokalklimatologische Interpretation von Thermalluftbildern", Heft4, Bildmessung und Luftbildwesen, pp. 152-158, 1975

Gates, D. M., "Infrared measurement of plant and animal surface temperature and their interpretation", Remote Sensing in Ecology, pp. 95-107, 1969

Goudriaan, J., "Crop micrometeorology - a simulation study", 1977, Wageningen - PUDOC report, Centre for Agricultural Publishing and Documentation

Haugen, D. A. (Ed.), "Workshop on Micrometeorology", Boston, American Meteorological Society, 1973, pp. 73-78

Higgins, G. J., Hilton, P. F., Shapiro, R., Touart, C. N., Wachtmann, R. F., "Operational Tactical Decision Aids (OTDA), Final Report, Volume II", 1987, AFGL-TR-87-0330, Hanscom AFB, MA

Hodges, D. B., Higgins, G. J., Hilton, P. F., Hood, R. E., Shapiro, R., Touart, C. N., Wachtmann, R. F., "Final Tactical Decision Aid (FTDA) for Infrared (8-12µm) Systems - Technical Background", 1983, AFGL-TR-83-0022, Hanscom AFB, MA

Hoyningen-Huene, J. v., Bramm, A., "Zur Bedeutung der Assimilationsmechanismen für den Energie- und Wasserhaushalt von Kulturpflanzen", Met. Rdsch., 34, pp. 167-178, 1981

Idso, S. B., Schmugge, T. J., Jackson, R. D., Reginato, R. J., "The Utility of Surface Temperature Measurements for Remote Sensing of Surface Soil Water Status", J. Geophys. Res., 80, pp. 3044-3043, 1976

Jaeger, L., "Radiation measurements of the Department of Meteorology of Freiburg University during the experimental stage of MESOKLIP", Met. Rdsch., 34, pp. 97-105, 1981

Leidner, L., Clement, D., Levin, I., Messner, H., "Variation of Surface Temperature Models under European Conditions", paper held at Ninth Annual EOSAEL/TWI conference, 29 November - 1 December 1988, Phys. Sci. Lab., Las Cruces, NM

Linacre, E. T., "Further notes on a feature of leaf and surface temperatures", Archiv f. Met. Geophys. und Bioklimat., B 15, pp. 422-436, 1967

Mansi, M. V., Walls, I. A., "Prediction of Temeprature Difference in Background Scenery", in "Passive Infrared Systems and Technology, SPIE Proceedings Vol. 807, pp. 61-68, 1987

Monsi, M., Saeki, T., "Über den Lichtfaktor in den Pflanzengesellschaften und seine Bedeutung für die Stoffproduktion", Jap. J. Bot., 14, pp. 22-52, 1953

Monteith, J. L. (Ed.), "Vegetation and the Atmosphere Vol I", 1975, London, Academic Press

Nickerson, E. C., Smiley, V. E., "Surface energy budget parameterizations for urban scale models", J. Appl. Met., 14, pp. 297-300, 1975

Pearman, G. I., Weaver, H. L., Tanner, C. B., "Boundary layer heat transfer coefficients under field conditions", Agric. Met., 10, pp. 83-92, 1972

Raschke, K., "Über die physikalischen Beziehungen zwischen Wärmeübergangszahl, Strahlungsaustausch, Temperatur und Transpiration eines Blatts", Planta, 48, pp. 200-238, 1956

Shapiro, R., "Solar Radiative Flux Calculations from Standard Surface Meteorological Observations", 1982, AFGL-TR-82-0039, Hanscom AFB, MA

Shawcroft, R. W., Lemon, E. R., Allen, L. H., Stewart, D. W., Jensen, S. E., "The soil-plant-atmosphere model and some of its predictions", Agric. Met., 14, pp. 287-307, 1974

Waggoner, P. E., Reifsnyder, W. E., "Simulation of the temperature, humidity and evaporation profile in a leaf canopy", J. Appl. Met., 7, pp. 400-409, 1968

Wollenweber, F. G., "Die Verdunstung natürlicher Oberflächen und ihre Modellierung mit einem einfachen Bestandsklimamodell", in Freiburger Geographische Hefte, Heft 26, International Symposium on Urban and Local Climatology Freiburg(Br.), February 20-21, 1986

DISCUSSION

F. CHRISTOPHE, FR

Do you expect that a model derived from yours to be suitable at mm-wavelengths; that is, for predicting contrast for a millimeter wave imaging radiometer?

AUTHOR'S REPLY

The outlined model is able to predict surface temperatures of vegetation canopies. To determine mm-wavelength radiances from these results, it would be necessary to know emissivities of vegetation in this wavelength region. In the IR-region, emissivities are very close to unity, which is not the case for mm-wavelengths. Consequently, the problem is more complicated there.

J. HANCOCK, US

The reason we are dropping empirical models and are developing first principles models is that we felt the empirical models were not producing surface temperature forecasts accurately enough. We found it very difficult to adjust the regression coefficients to improve model prediction. When first principles models performance is insufficient, we are able to physically explain the reasons for prediction error, and adjust modeling procedures to correct these deficiencies.

AUTHOR'S REPLY

I do not want to comment further on Major Hancock's statement, since it gives a concise summary of a major result of this paper.

Remote Sensing of the Aerosol Scattering Coefficient with a Multi-Field-of-View Lidar

Luc R. Bissonnette and Daniel L. Hutt
DREV-Defence Research Establishment Valcartier
P.O. Box 8800,Courcelette, Québec, Canada , G0A 1R0

Abstract

A new lidar technique is proposed for the remote determination of the obscuration effects of natural and man-made aerosols. It is based on the simultaneous measurement of the lidar returns at different fields of view. By ratioing these returns, we eliminate the need for a backscatter-to-extinction relation which makes the inversion of the single-scattering lidar measurements subject to uncertainties. A system operating at 1.054 μm and fields of view of 5, 25, 50 and 75 mrad is described. Preliminary results are discussed.

1 Introduction

Many modern defence systems use electro-optic (EO) devices for surveillance, detection, identification, ranging, and guidance and control. These systems work well in good weather conditions but can be adversely affected in poor visibility created by natural aerosols or man-made obscurants. Therefore, field commanders need a continously updated assessment of the performance of their EO systems.

Numerous studies have been carried out of atmospheric propagation effects in the visible and the infrared. These have led to the development of complex models to calculate the atmospheric extinction coefficient as a function of wavelength and meteorological parameters. The most widely used models are LOWTRAN [1,2], FASCOD [3,4], and the EOSAEL [5] library of computer programs. These models are very useful for design purposes, for climatological studies, and for defining operational guidelines on the efficient use of EO weapon systems. However, they are of limited help for real-time, on-site deployment decisions. The physics of the aerosols which contribute significantly to the optical and infrared obscuration is simply too complex to be predictable with acceptable accuracy on the temporal and spatial scales relevant to field commanders. Real-time and *in situ* measurements are therefore necessary.

A useful real-time and *in situ* measurement is the visibility or meteorological range. In daytime and on land, very good estimates of the visibility can be obtained. However, at sea and at nighttime, it is either not available or unreliable. In any case, visibility alone is not sufficient as there is often little correlation between the extinction coefficients in the visible and the infrared bands. Therefore, some means of measuring directly the aerosol extinction coefficients at the wavelengths of interest must be found. The measured extinction coefficients along with the standard meteorological parameters could then be integrated into relevant Tactical Decision Aids to assist field commanders in making deployment decisions. For operational purposes, the measurements must be done remotely from a single station.

The remote measurement of the aorosol optical and infrared extinction coefficients is not yet a proven technology. The most promising approach is the lidar technique, or the inversion of the range-gated aerosol-backscattered signal from a short laser pulse. One difficulty with this technique is that the measured backscatter is a function of two unknowns: the backscatter and the extinction coefficients. In this paper, we briefly review the problem and propose a multi-field-of-view method that may contribute to resolving the aforementioned indeterminacy. We also describe a multi-field-of-view lidar instrument that has just been built and discuss some preliminary results.

2 Single-Scattering Lidar Inversion

For a monostatic and monochromatic system, the single-scattering lidar equation [6] is

$$P(R) = K\frac{\beta(R)}{R^2}\exp\left[-2\int_0^R \sigma(r)dr\right], \tag{1}$$

where $P(R)$ is the power received from range R, K is the system constant, and $\beta(R)$ and $\sigma(R)$ are the backscatter and extinction coefficients. Equation 1 relates two unknowns, $\beta(R)$ and $\sigma(R)$, to one measured quantity, $P(R)$. The most common approach to handling this problem is to assume a relation between β and σ of the form

$$\beta = C\sigma^k, \tag{2}$$

where C can be a function of R but where k is usually assumed constant. Equation 2 is substituted for β in eq. 1 and after differentiating with respect to R, the following solution is obtained [7]:

$$\sigma(R) = \frac{\left[\frac{R^2}{R_b^2}\frac{C(R_b)}{C(R)}\frac{P(R)}{P(R_b)}\right]^{1/k}}{\sigma_b^{-1} + \frac{2}{k}\int_R^{R_b}\left[\frac{r^2}{R_b^2}\frac{C(R_b)}{C(r)}\frac{P(r)}{P(R_b)}\right]^{1/k}dr}, \tag{3}$$

where R_b is the range at which the boundary value σ_b is specified. R_b can be greater or smaller than R. It is assumed in the following that $C(R)$ is known or can be determined.

Two principal difficulties are associated with the lidar solution given by eq. 3. First, eq. 3 assumes a single-valued relation between the backscatter and extinction coefficients. Pinnick *et al.* [8], Takamura *et al.* [9] and Evans [10] have shown through measurements and calculations that, although there are practical conditions for which eq. 2 is a reasonable approximation, β/σ is widely spread in many cases. The second difficulty is the need for a boundary value σ_b. This boundary value cannot be arbitrarily chosen, it must be consistent with the measured $P(R)$.

The problem of the boundary value has been dealt with in various ways. Klett [11] showed that integrating backward, *i.e.* specifying the boundary value at $R_b > R$, leads to a more stable solution, a solution that tends to the true solution with decreasing R despite errors on σ_b. Uthe *et al.* [12] used the total optical thickness measured by other means over a given range in lieu of a boundary value. Finally, Kunz [13] derived his boundary

value for slant path returns by first inverting a horizontal shot under the assumption of horizontal homogeneity.

Perhaps the most serious problem is the $\beta - \sigma$ relation. Without additional independent measurements, one cannot hope to achieve a fully unambiguous solution. This constitutes a constraint that could make the use of lidars impractical for true single-ended remote measurements. For example, Kunz [14] and Paulson and Powers [15] proposed a doubled-ended lidar technique that provides two independent measurements to solve for both β and σ. However, this requires setting up a range instrumented at both ends and is therefore inappropriate for true remote sensing.

In summary, the single-scattering lidar method can still provide a wealth of information on the optical and infrared properties of aerosols but quantitative results can be subject to relatively large errors due to uncertainties on the boundary value and the β/σ ratio. A theoretical analysis of the sensitivity of the lidar solution to these error sources is given in [16]. In what follows, we propose to measure simultaneously the lidar returns at different fields of view in the hope of gaining additional information to substitute for the assumption of a $\beta - \sigma$ relation. The advantage is that this would still be a single-ended instrument.

3 Multiple-Scattering Lidar Technique

Equation 1 is called the single-scattering lidar equation. It takes into account the radiation that has undergone only one scattering, *i.e.* the backscattering at range R. All the other scattering events occurring on both the forward and backward propagation legs are assumed to remove completely the radiation from the receiver field of view. In reality, forward-scattered radiation can reach the receiver, especially in the case of scattering aerosol phase functions peaked in the forward direction. It is the information carried by these multiple-scattering contributions that we hope to use to solve for the scattering coefficient independently of the backscatter coefficient.

A multiple-scattering lidar equation can be derived from the lateral diffusion model of [17]. However, we get more physical insight by deriving the equation from phenomenological arguments.

A first effect of scattering is that the aerosol volume responsible for the lidar backscatter from range R is illuminated not only by the unscattered reduced-intensity laser beam but by light that has been scattered forward along the path from the transmitter to range R. The incident power $P_i(R)$ illuminating a volume of unit length and cross-section determined by the receiver field of view θ is thus given by

$$P_i(R) \propto P_0 e^{-\eta_a - \eta^-} \left\{ e^{-\eta^+} + \left[1 - e^{-\eta^+} \right] f(R^2\theta^2/W^2) \right\}, \tag{4}$$

where P_0 is the laser average power over its pulse length, $\eta_a = \int_0^R \sigma_a dr$ with σ_a the absorption coefficient, $\eta^\pm = \int_0^R \sigma^\pm dr$ with $\sigma^\pm$ the backward- and forward-scattering coefficients obtained by integrating the angular scattering function over the backward and forward hemispheres, f is the fraction of scattered radiation contained within the field of view θ, and $W^2(R, \eta^+, \phi)$ measures the lateral spreading of the multiply forward-scattered radiation with ϕ a path weighted average of the forward-scattering angle. The functions f and W^2 are yet unspecified. The first term represents the contribution from the unscattered beam and the second term, the contribution from forward scattering.

Similarly, part of the radiation that is backscattered at range R reaches the receiver without being scattered and part is contributed by one or many scattering events. Thus, the power $P(R)$ received from range R is approximated by

$$P(R) \propto \frac{\sigma^-(R)P_i(R)e^{-\eta_a-\eta^-}}{G}\left\{e^{-\eta^+} + \left[1 - e^{-\eta^+}\right]g(\theta^2/\varphi^2)\right\}, \tag{5}$$

where σ^- is the backscatter coefficient, $g(\theta^2/\varphi^2)$ is the fraction of the radiation forward-scattered on the return propagation leg that is captured by the receiver field of view, φ is a path weighted average of the forward-scattering angle, different in principle from ϕ, and G is a geometric factor that does not need to be put in explicit form for the present application. The function f will be specified later.

Combining eqs. 4 and 5, we have

$$P(R) \propto \frac{\sigma^-(R)e^{-2\eta}}{G}\left[1 + \left(e^{\eta^+} - 1\right)f\right]\left[1 + \left(e^{\eta^+} - 1\right)g\right], \tag{6}$$

where $\eta = \eta_a + \eta^- + \eta^+$ is the optical depth. Equation 6 neglects the stretching of the laser pulse due to multiple scattering. Pulse stretching blurs the range resolution but negligibly so for original pulse lengths of the order of tens of ns and for fields of view smaller than a few degrees.

Expressions for the functions f and g can be obtained from the model of [17]. They are complicated integral equations but useful asymptotic limits are easily derived. f and g are asymptotic to their respective arguments when the latter tend to zero and to unity when they tend to infinity. For the present application, we postulate the following simple formulas to connect the two limits:

$$f(x) = \frac{x}{1+x}, \tag{7}$$

$$g(y) = \frac{y}{1+y}, \tag{8}$$

where $x = R^2\theta^2/W^2$ and $y = \theta^2/\varphi^2$. The function W^2 can also be derived in integral form by use of the model solutions of [17] but the following expression constitutes a useful approximation:

$$W^2 \simeq \kappa\phi R^2 e^{\eta^+}, \tag{9}$$

where κ is an empirical constant with the dimension of an angle. Assuming that $\varphi \simeq \phi$ and substituting eqs. 7-9 in eq. 6, we have

$$P(R) \propto \frac{\sigma^- e^{-2\eta}}{G}e^{\eta^+}\left[\frac{\theta^2 + \kappa\phi}{\theta^2 + \kappa\phi e^{\eta^+}}\right]\left[\frac{\phi^2 + \theta^2 e^{\eta^+}}{\phi^2 + \theta^2}\right]. \tag{10}$$

Both approximations $W^2 \simeq \kappa\phi R^2 \exp(\eta^+)$ and $\varphi \simeq \phi$ will require validation.

For the multiple-scattering technique lidar, we propose to make simultaneous return measurements at different fields of view θ. By ratioing the measured signals, the unknown functions σ^- and G drop out and we are left with

$$\frac{P(R,\theta)}{P(R,\theta_1)} = \left(\frac{\theta^2 + \kappa\phi}{\theta_1^2 + \kappa\phi}\right)\left(\frac{\phi^2 + \theta_1^2}{\phi^2 + \theta^2}\right)\left(\frac{\theta_1^2 + \kappa\phi e^{\eta^+}}{\theta^2 + \kappa\phi e^{\eta^+}}\right)\left(\frac{\phi^2 + \theta^2 e^{\eta^+}}{\phi^2 + \theta_1^2 e^{\eta^+}}\right). \tag{11}$$

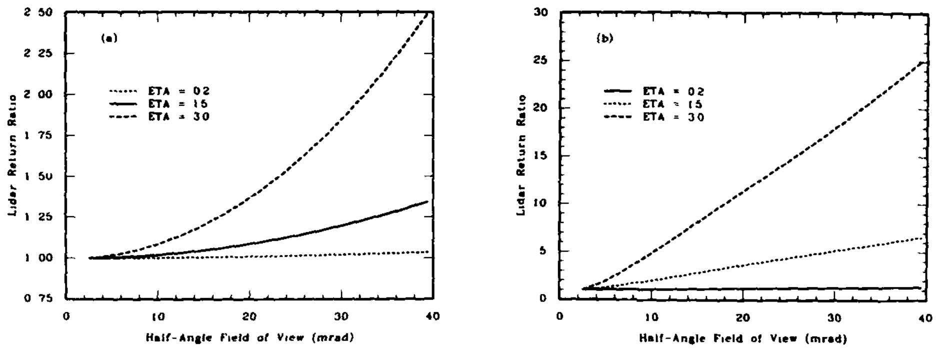


Figure 1: Curves of the lidar return ratio versus field of view (half angle) calculated with eq. 13 for different η^+ values, $\kappa = 0.08$: a, for $\phi = 0.150$; b, for $\phi = 0.015$.

We use as the reference signal $P(R, \theta_1)$ the return measured with the smallest field of view. By design, θ_1 is chosen to be slightly greater than the laser beam divergence to also have as a special case the single-scattering lidar configuration. Hence, θ_1 is generally of the order of a few mrad and, for most cases, the following conditions should be satisfied:

$$\theta_1^2 \ll \phi^2;\ \kappa\phi, \tag{12}$$

except near a cloud boundary where $P(R,\theta)/P(R,\theta_1) \rightarrow 1$. Thus, in regions where $P(R,\theta)/P(R,\theta_1)$ is sufficiently different from unity for the multiple-scattering technique to be applicable, eq. 11 can be simplified as follows:

$$\frac{P(R,\theta)}{P(R,\theta_1)} = \left(\frac{\theta^2 + \kappa\phi}{\theta^2 + \phi^2}\right)\left(\frac{\phi^2 e^{\eta^+}}{\phi^2 + \theta_1^2 e^{\eta^+}}\right)\left(\frac{\phi^2 + \theta^2 e^{\eta^+}}{\theta^2 + \kappa\phi e^{\eta^+}}\right). \tag{13}$$

Equation 13 shows that the measured ratio is function of two unknown quantities: η^+, the scattering optical thickness; and ϕ, the average forward-scattering angle. If the ratio is measured at two or more field-of-view angles θ that are not small compared with ϕ, we can in principle solve for both η^+ and ϕ. The equations are algebraic equations that do not require boundary values and the solution at each point R is independent of the solutions at other points. Hence, errors do not propagate. For example, the expected larger errors at small η^+ would not affect the results at greater optical depths.

Figure 1 shows examples of how the $P(R,\theta)/P(R,\theta_1)$ given by eq. 13 varies with θ for selected values of η^+ and ϕ. For a given ϕ, it is obvious that the ratio is a resolvable function of η^+ but it may happen that different sets of ϕ and η^+ produce ratios of similar magnitude. One such case is illustrated in Fig. 2. The two lidar return ratios vary over a comparable range of values for sets of η^+ and ϕ that differ by a factor of 7.5 and 10, respectively. However, the difference in the θ-dependence is such that solving for both η^+ and ϕ is possible. Hence, the proposed inversion method appears feasible but the errors are expected to grow with the relative magnitude of ϕ compared with the θ's.

By differentiating η^+ with respect to range R, we obtain the forward-scattering coefficient σ^+. For aerosols with an albedo close to unity and a phase function reasonably peaked in the forward direction, σ^+ is within 10-15 % of the extinction coefficient. These conditions are well verified in atmospheric aerosols for wavelengths in the near infrared.

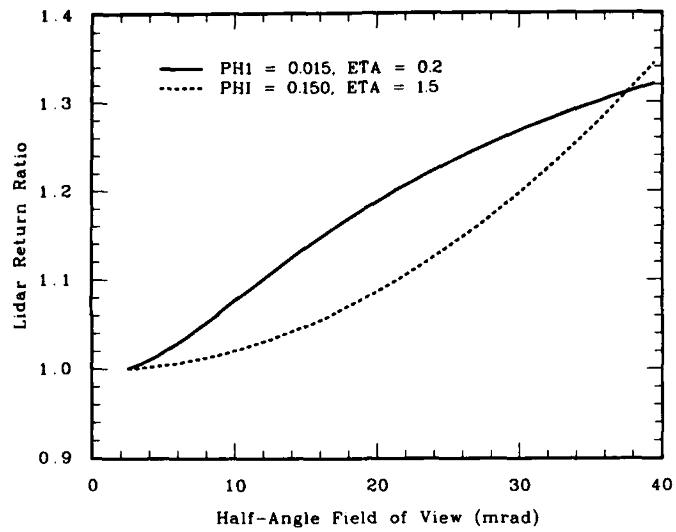


Figure 2: Comparison of calculated curves of the lidar return ratio versus field of view (half angle) for different sets of η^+ and ϕ.

Since the method also gives the average forward-scattering angle, it is hoped that this additional information can be applied to estimate the size of the aerosol particles which could then be used to extrapolate the measured scattering coefficient to other wavelengths.

In summary, the multiple-scattering technique lidar derives the forward-scattering coefficient from the information contained in the angular dependence of the received lidar returns. The solution method requires neither a boundary value nor a backscatter-to-extinction relation, and still relies on single-ended remote measurements.

4 Multi-Field-of-View Lidar

An instrument was built to test the multiple-scattering method discussed in the preceding section. An overall view of the lidar transceiver is shown in the photograph of Fig. 3. The laser transmitter and the telescope receiver are mounted on a scanner that can be positioned between $-5°$ and $+90°$ in elevation, and $-45°$ and $+45°$ in azimuth. The source is a pulsed Q-switched Nd:Glass laser at 1.054 μm. The laser has a maximum output energy of 1.5 J/pulse at its recommended maximum firing rate of 1 Hz and the pulse length is 25 ns. The coaxial radar shown in Fig. 3 is used for eye-safety reason, it will automatically close the laser shutter if any object enters the beam within the non-eye-safe range.

The novel feature of this instrument is the receiver that allows simultaneous lidar return measurements at four fields of view. The receiver main elements are a custom designed detector and a custom designed lens. The detector is a special silicon PIN diode. It is made up of four concentric active areas which are electrically insulated from each other. The outside diameters of the active areas are respectively 1, 5, 10 and 15 mm. Combined with the focal length of 202.46 mm of the telescope lens, these give nominal fields of view of 5, 25, 50 and 75 mrad. The diode has a response of 0.35 A/W at 1.054

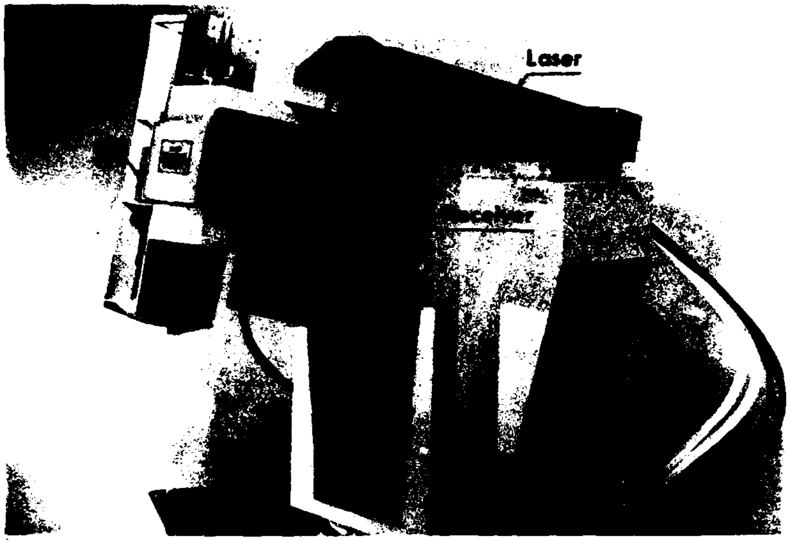


Figure 3: Photograph of the lidar transceiver.

μm and a dark current of less than 1 μA at its normal operating bias of 80 V. Each active area has a rise time of less than 35 ns. The uniformity of the diode response over its entire surface is better than 5 %. The separation between the active areas is 0.127-mm wide.

The main element of the telescope receiver is a custom designed 152.4-mm diameter f/1.33 lens assembly. Its blur circle over the whole area of the detector in the focal plane is smaller than the inter-element separation. A 850-nm cutoff filter in mounted in front of the telescope housing. All lens surfaces, including the cutoff filter, are anti-reflection coated.

The voltage output of each active area of the detector diode is conditionned by a logarithmic amplifier before digitization in the waveform recorder of the data acquisition system. These logarithmic amplifiers are used to compress the output signal of the PIN diode elements to make it compatible with the dynamic range of the waveform recorders. The input range is 20 μV to 3 V and the output swing is 2 V. The large signal response characteristics are 16-ns rise time and 25-ns fall time. Figure 4 is a photograph showing the detector and the logarithmic amplifiers assembled in a single module.

The laser source is mounted on top of the telescope housing as illustrated in Fig. 3. The beam is steered by two mirrors to bring its axis coincident with the telescope axis. A schematic of the optical arrangement is shown in Fig. 5. Mirror M1 is adjustable. Mirror M2 is elliptical with a minor axis diameter of 25.4 mm. It is set at a fixed angle of 45° but rotation about the telescope axis is possible. With the adjustments of M1 and M2, the laser beam can be centered and aligned on the telescope axis. This condition is essential for the proposed multiple-scattering method. The outgoing unscattered laser beam must either not come into the field of view of the outer detector rings or, if it does, it must be exactly centered. In the present instrument, the beam divergence is 4 mrad and the central field of view, 5 mrad. Hence, the alignment must be better than 1 mrad.

The four logarithmically amplified signals are simultaneously digitized at a rate of 32 MHz. All the data acquisition is controlled by software that also provides rapid dis-

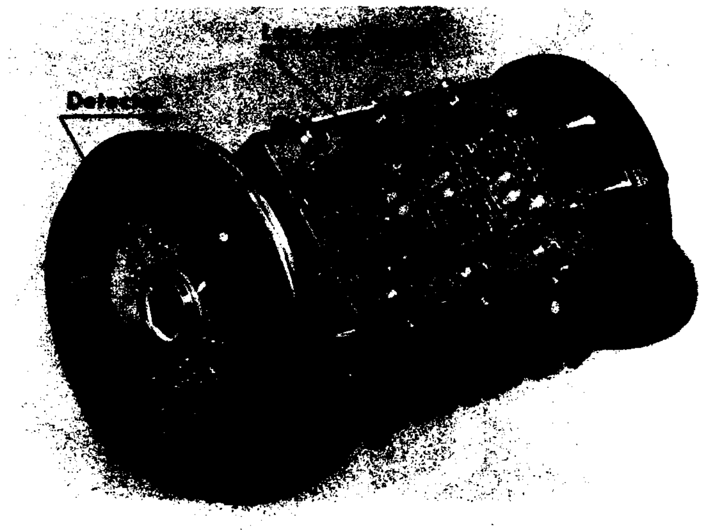


Figure 4: Photograph of the lidar detector module.

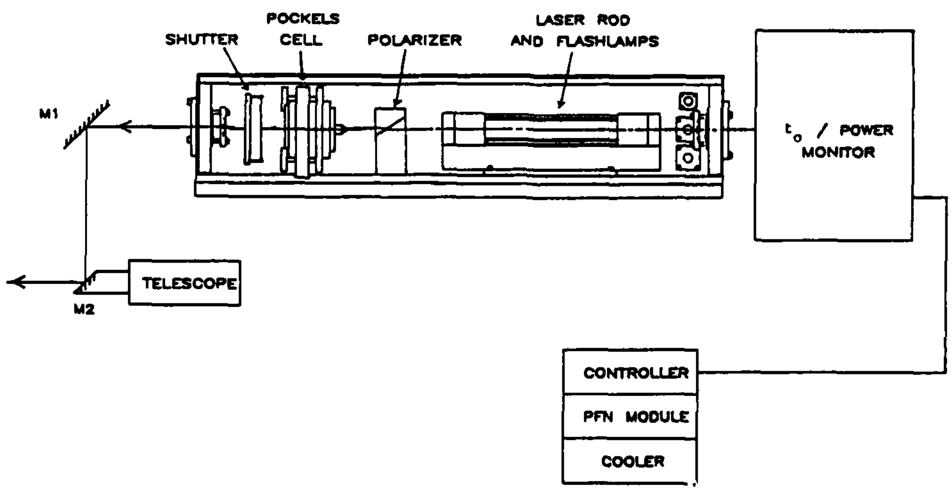


Figure 5: Diagram of the lidar transmitter.

LASER		RECEIVER		PROCESSING	
Wavelength	1.054 μm	Diameter	152 mm	Log Amp.	>4 dec.
Pulse Energy	1.5 J	Focal Length	202 mm	Dig. Rate	32 MHz
Pulse Length	25 ns	Det. Rings	1,5,10,15 mm	Max. Length	64 μs
Divergence	4 mrad	FOVs	5,25,50,75 mrad	of Memorized	or
Rep. Rate	< 1 Hz			Lidar Returns	9.6 km

Table 1: Multi-Field-of-View Lidar Characteristics.

play for monitoring the experiment in real time. Table 1 summarizes the main system characteristics.

5 Measurements

Preliminary lidar return measurements are plotted in Figs. 6a and 6b. These results are only partial. On the one hand, the fourth detector ring does not function properly; the signal does not recover rapidly enough after the initial hit by the radiation scattered off the optical surfaces. The fall time is greater than the 64 μs of the data record. This effect is attributed to the noncircular detector area that extends beyond the fourth ring element. Since this sensitive area is not biased, the charges produced at $t = 0$ are not swept away. Better optical shielding will be put in place to try to remedy this problem and provision is being made to rebuild the detector with a guard ring. On the other hand, our alignment procedure turned out to be not precise enough to ensure that the single-scattering return is unambiguously centered on the inner 5-mrad field of view. This is because the beam divergence is close to 5 mrad and that the beam profile at the time of the measurements reported in Figs. 6 was highly non-uniform with most of the energy concentrated in two hot spots near the edge of the beam. Further adjustments/improvements will be necessary to achieve a more uniform beam profile and reduce the telescope/laser alignment errors. Consequently, in Figs. 6, we show the returns from only two fields of view: 25 and 50 mrad. The 25- and 50-mrad curves were obtained by adding together the returns from the detector elements 1 and 2, and 1, 2 and 3, respectively.

Figures 6a and 6b show typical returns from rain clouds. The time separation between these two measurements is just over 6 minutes which illustrates that the cloud structure can change quite rapidly. Of special interest to the present application is the ratio of the 50- to the 25-mrad returns. This ratio corresponding to the measurements of Figs. 6a and 6b is plotted as a function of range in Figs. 7a and 7b, respectively. At first, there is no measurable signal from the outer detector element and the ratio remains unity. With further penetration, the multiple-scattering contributions begin to appear and the ratio starts to increase slowly up to the base of the cloud layers. The backscatter in this region is from a mixture of light rain and haze. At the base of the cloud layers, there is a slight drop followed by a rapid increase. It is worth noting that this rapid increase is somewhat delayed compared with that of the corresponding lidar returns. What we can conclude from these results is that the multiple-scattering contributions are measurable, and that they are of magnitude comparable with the calculations based on the simple model of eq. 13 as shown by the curves of Figs. 1a and 1b. Hence, although the measurements to date are limited to two fields of view, which is not sufficient to test the proposed multiple-scattering inversion, the results of Figs. 6 and 7 indicate that the technique appears

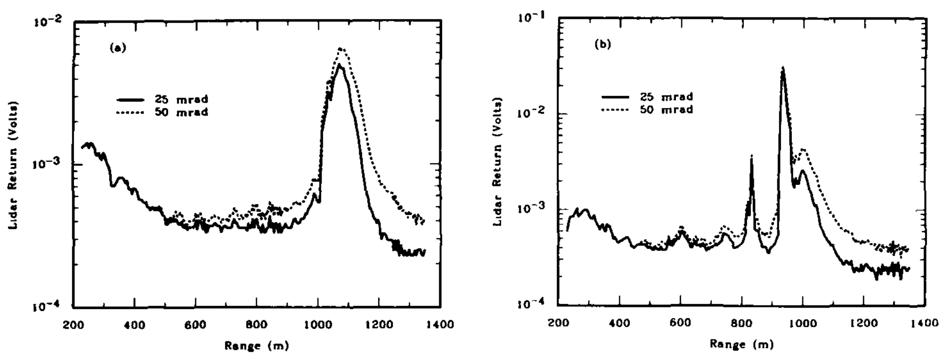


Figure 6: Simultaneous 25- and 50-mrad field-of-view lidar returns from rain clouds at 26-degree elevation: a, at 9:21:42 hrs; b, at 9:28:08 hrs.

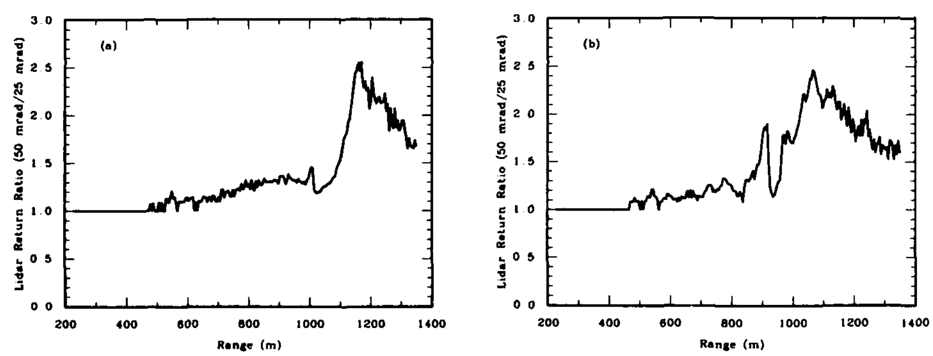


Figure 7: Measured ratio of the 50- to the 25-mrad lidar returns of Figs. 6a and 6b respectively.

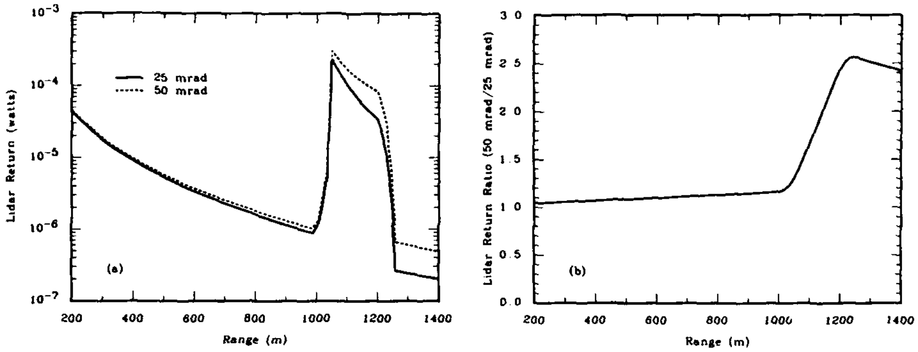


Figure 8: Calculated 25- and 50-mrad lidar returns (a), and their ratio (b), for conditions given in the main text.

feasible and is worth investigating.

To verify how modeling compares with measurements, we ran a computer simulation using the model of [17]. We assumed a 250-m thick cloud layer with an extinction coefficient of 10/km and located 1 km away from the lidar. Below the cloud, we assumed uniform light rain with an extinction value of 0.5/km and beyond, clear atmosphere with extinction equal to 0.1/km. This represents a simplified cloud/atmosphere picture estimated from the measured curves of Fig. 6a. The pertinent scattering parameters were derived from the EOSAEL [5] phase function data file. The calculated lidar returns and lidar return ratio are plotted in Figs. 8a and 8b, respectively. The resemblance with the experimental curves of Figs. 6a and 7a is very close. Since the proposed multiple-scattering inversion method is based on the same model as that used in the calculations of Figs. 8, these results further indicate that useful information should indeed be derivable from multi-field-of-view lidar measurements.

6 Conclusion

A multiple-scattering lidar technique for the determination of the atmospheric aerosol scattering coefficient was defined. The method requires the simultaneous measurement of lidar returns at three or more fields of view. The principal advantage is that neither a boundary value nor a backscatter-to-extinction relation are needed. To verify the method, a four-field-of-view system operating at 1.054 μm was built. Because of difficulties with two field-of-view channels, results to date are only preliminary but, nevertheless, they indicate that the technique is feasible. In particular, it was shown that the multiple-scattering contributions to the backscatter are not only measurable but of order of magnitude in agreement with model predictions. Complete validation will require an extended experimental program with careful and detailed monitoring of the atmospheric aerosol properties. The simple heuristic model of this paper will probably need adjustments, especially eqs. 7-9 and the approximation $\varphi \simeq \phi$. In addition to the proposed inversion application, the new lidar will certainly provide useful information on the effect of multiple scattering on lidar measurements.

References

[1] McClatchey,R.A., Fenn,R.W., Selby,J.E.A., Volz,F.E. ang Garing,J.S., "Optical Properties of the Atmosphere", AFCRL-72-0497, 1972.

[2] Kneizys,F.X., Shettle,E.P., Gallery,W.O., Chetwynd,J.H., Abreu,L.W., Selby,J.E.A., Clough,S.A. and Fenn,R.W., "Atmospheric Transmittance/Radiance: Computer Code LOWTRAN6", AFGL-TR-83-0187, 1983.

[3] Clough,S.A., Kneizys,F.X., Rothman,L.S. and Gallery,W.O., "Atmospheric Spectral Transmittance and Radiance: FASCOD1B", Proc. Soc. Photo-Opt. Instrum. Eng., Vol. 277, 1981.

[4] Ridgway,W.L., Moose,R.A. and Cogley,A.C., "Atmospheric Transmittance/Radiance Computer Code FASCOD2", AFGL-TR-82-0392, 1982.

[5] Duncan.L.D. and Shirkey,R.C., "EOSAEL 82- A Library of Battlefield Obscuration Models", Opt. Eng, Vol. 22, pp. 20-23, 1983.

[6] Carswell,A.I., "Lidar Measurements of the Atmosphere", Can. J. of Physics, Vol. 61, pp. 378-395, 1983.

[7] Klett,J.D., "Lidar Inversion with Variable Backscatter/Extinction Ratios", Appl. Opt., Vol. 24, pp. 1638-1643, 1985.

[8] Pinnick,R.G., Jennings,S.G., Chýlek,Petr, Ham,C. and Grandy,W.T. Jr, "Backscatter and Extinction in Water Clouds", J. Geo. Res., Vol. 88, pp. 6787- 6796, 1983.

[9] Takamura,T. and Yasuhiro,S., "Ratio of Aerosol Backscatter to Extinction Coefficients as Determined from Angular Scattering Measurements for Use in Atmospheric Lidar Applications", Opt. and Quant. Electronics, Vol. 19, pp. 293-302, 1987.

[10] Evans,B.T.N., "Sensitivity of the Backscatter/Extinction Ratio to Changes in Aerosol Properties: Implications for Lidar", Appl. Opt., Vol. 27, pp. 3299-3305, 1988.

[11] Klett,J.D., "Stable Analytical Inversion Solution for Processing Lidar Returns", Appl. Opt., Vol. 20, pp. 211-220, 1981.

[12] Uthe,E.E. and Livingston,J.M., "Lidar Extinction Methods Applied to Observations of Obscurant Events", Appl. Opt., Vol. 25, pp. 678-684, 1986.

[13] Kunz,G.J., "Vertical Atmospheric Profiles Measured with Lidar", Appl. Opt., Vol. 22, pp.1955-1957, 1983.

[14] Kunz,G.J., "Bipath Method as a Way to Measure the Spatial Backscatter and Extinction Coefficients with Lidar", Appl. Opt., Vol. 26, pp. 794-795, 1987.

[15] Paulson,M.R. and Powers,W.J., "Description of a Two-Lidar Technique for Determining Aerosol Extinction Coefficient Profiles", NOSC-TN-1449, 1986.

[16] Bissonnette,L.R., "Sensitivity Analysis of Lidar Inversion Algorithms", Appl. Opt., Vol. 25, pp. 2122-2125, 1986.

[17] Bissonnette,L.R., "Multiscattering Model for Propagation of Narrow Light Beams in Aerosol Media", Appl. Opt., Vol. 27, pp. 2478-2484, 1988.

DISCUSSION

S. CLIFFORD, US

What are the effects of turbulence on your lidar measurements of multiple scattering from clouds?

AUTHOR'S REPLY

We routinely observe small-structure fluctuations in the return signal. These are due to fluctuations in aerosol concentration and size distribution which are certainly related to atmospheric inhomogeneities or turbulences. We have not yet investigated these effects which could be important for scales of the order of the field-of-view aperture. My guess is that the method will yield, at best, an average value over the field-of-view aperture.

Temporal fluctuations have no effect since the duration of the pulse is much shorter than any characteristic time of atmopspheric turbulence.

H. HUGHES, US

Could you clarify two points for me? The first concerns the measurements that were presented for the 25 mrad and 50 mrad fields-of-view. There the power ratios did not exceed unity until near the cloud base. While the top of the cloud was not identified in the measurements, in the attempts at modeling the returns, you showed power ratios exceeding unity above the cloud top where the clear air extinction coefficients corresponded to nearly 40 km visibility. Is this feature an artifact of the EOSAEL phase functions that were used in the calculations? Second, could you comment on how useful this technique might be in situations other than fogs or clouds for visibilities greater than 2-4 km which are of primary interest to the Navy?

AUTHOR'S REPLY

The feature you are referring to is not an artifact. The aerosol beyond the cloud is illuminated by the unscattered beam as well as by radiation forward-scattered by the cloud. The latter contribution is spread angularly and therefore can cause a measurable return for a field-of-view greater than 25 mrad even though the aerosol concentration in this region is very small and would not by itself give rise to significant multiple scattering.

At the wavelength of 1.05 microns, computer simulations indicate that the technique should apply for visibilities in the range of 2-8 km. It is possible, however, that a different set of field-of-view values would be more appropriate, probably smaller than for the current system.

PHYSICAL MODELS FOR AEROSOL IN THE MARINE MIXED-LAYER

by
Gerrit de Leeuw,* and Kenneth L. Davidson
Naval Postgraduate School, Department of Meteorology
Monterey, California 93943, U.S.A.

*permanent address: Physics and Electronics Laboratory TNO
P.O. Box 96864, 2509 JG The Hague, The Netherlands

Stuart G. Gathman
Naval Research Laboratory, Code 4117
Washington D.C., 20375, U.S.A.

R.V. Noonkester+

Naval Ocean Systems Center
San Diego, California 92152-5000, U.S.A.

+present address: Innovative Meteorology,
4247 Mnt. Henry Ave, San Diego, CA 92117, USA

SUMMARY

A model is presented to calculate the vertical variation of aerosol extinction coefficients throughout the marine atmospheric boundary layer. It is referred to as the Naval Oceanic Vertical Aerosol Model (NOVAM). NOVAM is a combination of empirical and physical models, formulated to describe the often observed non-uniform, but also non-logarithmic, profiles. The physical model is based on the dynamical processes affecting the production, mixing, deposition and size of the aerosol within the marine atmosphere. A status report is presented including a critical evaluation.

1. INTRODUCTION

For the assessment of the application of electro-optical (EO) systems for vertical and slant path observations, the height variation of electromagnetic scattering and absorption at wavelengths in the visible to the infrared is of considerable interest. In the evaluations of EO propagation characteristics problems arise because existing empirically derived expressions for aerosol scattering and absorption contributions to extinction were formulated for single levels. The Naval Aerosol Model (NAM)[1] as found in Lowtran VI is an example of this limitation. All data used in the development of NAM was derived from deck level measurements and no real provision was made for vertical structure in the aerosol concentration. Variations in the vertical may be very large, however.

When vertical structure is required for slant path calculations, the surface aerosol concentrations need to be extended to higher levels. This may be based on empirical models or on physical arguments. A usual approach in existing empirical models is to assume a logarithmic decrease with height using effective scaling heights.[2] Observed profiles are often non-logarithmic, however. Therefore extra information available from observed meteorological profiles should be used to take into account the physical processes which influence the vertical aerosol structure and which are thought to be responsible for the observed variety of profiles. Existing empirical models do not allow for the use of this extra information.

In this paper we describe an approach being formulated to put vertical structure into the extinction prediction using a mixture of empirical and physical models which describe the aerosol dynamical behavior. Prediction in this context does not imply prediction in time but rather an estimate of optical extinction given a set of atmospheric parameters which can be used with the empirical-dynamical model. The model is referred to as the Naval Oceanic Vertical Aerosol Model (NOVAM).[3]

The model for the structure in extinction was designed to describe non-uniform but also non-logarithmic aerosol distributions which are observed to exist throughout the marine atmospheric boundary layer. It is restricted to the marine atmosphere, hence the designation Oceanic in its title. The differences between this model and land-based models are the marine type of scaling used for the turbulent controlled processes near the sea surface, and the determination of the surface concentrations with NAM. The structure is a function of turbulent controlled processes and of the growth of the particles due to height varying relative humidity. The turbulent processes produce, deposit and mix the aerosol and also determine the depth of the mixed layer itself.

The following aspects of the multi-component model are addressed. The physical background of the turbulent controlled processes and of the growth features caused by relative humidity effects are presented in section 2, as well as the model used to predict

extinctions found under solid cloud decks. The model architecture and considerations of the physical constraints as they are treated in NOVAM are presented in section 3. Examples of calculated profiles are presented and compared with observed profiles. A critical evaluation of the several crucial parts of NOVAM with reference to its intended use appears in section 4. Finally, conclusions on the present status and future of the approach and model will be given.

2. PHYSICAL BACKGROUND

2.1 TURBULENT TRANSPORT IN THE MARINE ATMOSPHERIC BOUNDARY LAYER

The concentration of aerosols at various levels in the marine boundary layer is determined by a number of inter-dependent complex processes. Multi-variable models of this behavior are still in a rather crude state of development.

In our empirical-dynamical approach the starting point for modeling aerosol properties is the continuity equation including source, sink, vertical transport and 'horizontal advection' terms for the domain. Since the marine boundary layer is of limited vertical extent, both the surface and the top of it are potential source or sink regions.

As they are presently used in NOVAM, the dynamic equations neglect advection, the effects of which are included through the air mass parameter. Thus, the sources and sinks for the aerosol particles in the boundary layer are by transfer through the sea surface or by entrainment and gravitational fallout from the non-turbulent troposphere immediately above the marine boundary layer.

The vertical mixing of aerosols throughout the boundary layer is determined by the turbulent transport processes, which in turn are influenced by the relative humidity. The simplest case is the mid-latitude (as opposed to tropical) boundary layer with a strong inversion, which is well-mixed. When weak cumulus convection is present, a two layer model must be used to describe the aerosol structure.

In the simplest well-mixed case four scaling regimes exist within the marine boundary layer. These regimes are differentiated by the relative dominance of the different processes found within them. These are designated (see Figure 1) as the free troposphere above the mixed-layer (p), the mixed layer (f), the turbulent surface layer (c) and the diffusion sublayer (d).

The nature of the various atmospheric transfer processes permits us to identify certain height regimes where the analysis can be simplified by scaling arguments. For example, near the surface (within 10 meters of the ocean) the particle flux is generally considered independent of height. A

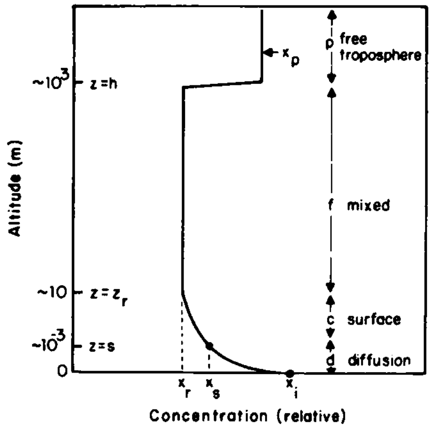


Figure 1. Schematic diagram of atmospheric scaling regimes (nonlinear scales).

thorough consideration of air-sea particulate transfer processes by Fairall and Larsen[4] addressed the relative importance of turbulent and diffusive transport mechanisms in this so-called constant flux layer and the diffusion dominated sublayer. Using a standard micrometeorological formalism, the surface source and sink properties can be described in a surface layer scaling context.

The mixed-layer constitutes about 90% of the boundary layer. Models based on its special properties are usually referred to as mixed-layer models. The classic mixed-layer model[5] is considered to be applicable to the mid-latitude marine regime where mixing in the boundary layer is dominated by reasonably homogeneous turbulence produced by surface shear and/or convection generated by warm water or cloud top radiative cooling. The mixed-layer model is one of the simplest because it ignores the details of the vertical transport processes by assuming that the turbulence is strong enough to maintain a well-mixed boundary layer. This implies that the fluxes in the boundary layer have a linear dependence on height and that we need only to specify the value of the flux at the bottom and top of the boundary layer.

The definition of the mixed-layer implies that particles of less than 30 μm radius are expected to obey mixed layer scaling[6] which is usually taken to mean the absence of a vertical gradient. Since the mixed layer formulation only requires that the gradient be constant with respect to time, clearly a constant vertical gradient is permissible. Davidson and Fairall,[7] using physical arguments of Wyngaard and Brost,[8] show that a mixed layer gradient for a surface generated aerosol component (e.g. sea salt) would be given by

$$\frac{\partial X_{sr}}{\partial z} = \frac{-1.5}{hw_*}(S_r - V_d X_{sr} + 2.5\, W_e X_{sr}) \quad (1)$$

where X_{sr} is the concentration of the sea-salt aerosol in the mixed-layer at height z, S_r is the surface flux, V_d is the effective fall velocity, W_e is the entrainment rate, h is the height of the boundary layer, and w_* is the convective scaling velocity.

Gradients predicted by Eq. (1) would be dependent on particle size. With representative values for the scaling parameters, the height variations for the very small particles can usually be neglected under typical conditions. This is generally not true for larger particles. It is also important to note that the role of relative humidity, which affects the gradient through both V_d and X_{sr}, has not been considered. This will be discussed in more detail in section 2.2.

Another climate regime is also globally important. This regime, which is visually characterized by "fair weather" or scattered cumulus clouds, is common over the ocean in the trade-wind latitudes. Physically, the presence of the cumulus towers significantly modifies the transport properties of the boundary layer. The cumulus towers dominate the upward transport of moisture, heat, and aerosols. This upward transport, which is confined to narrow columns that represent only a few per cent of the horizontal area, is balanced, in part, by a much more broadly spread downward transport (between cloud subsidence). A trade-wind equivalent to the mid-latitude mixed-layer model was developed by Albrecht.[9] Davidson and Fairall[7] describe the application to aerosols.

2.2 AEROSOL HUMIDITY EFFECTS

The marine aerosol consists in large part of hygroscopic particles, the size of which varies by evaporation and condensation, in response to changes in the relative humidity. In the mixed-layer the relative humidity varies with height and the sizes of the dispersing particles change accordingly. In NOVAM, the modal[3] aerosol concentration profile is determined for the size distribution at 80% relative humidity. For simplicity the humidity growth effects are only taken into account to adjust sizes and refractive index to derive the extinction coefficients, but not to alter the modal profile concentrations.

2.3 EXTINCTION IN MARINE STRATUS CLOUDS

The model used in NOVAM to calculate extinction in marine stratus clouds is distinctly different from the physical profile models described above. The stratus case bypasses estimates of aerosol entrainment, generation or deposition rates and is based only on the physics of aerosol growth with changes in relative humidity.

It was developed from detailed measurements in marine stratus cloud layers when the surface wind was low.[10] This limits application of the stratus model to cases when low level mixing is present and an inversion exists below 3 km, the cloud cover is greater than 0.8 and the wind speed does not exceed 5 m/s.

The extinction properties are determined using Fitzgerald's[11] approximation formulas which apply at wavelengths in the infrared between 1 and 11 μm, as compared to the wavelength of 0.2-40 μm for the other categories. This is a major limitation of the marine stratus model.

3. THE NAVAL OCEANIC VERTICAL AEROSOL MODEL (NOVAM)

3.1 INTENDED USE OF NOVAM

NOVAM was formulated to estimate the effect of the vertical variation of the aerosol concentration on slant path extinction. It is intended to be used with an equilibrium surface layer aerosol model such as NAM.[1] As such NOVAM is an extension of NAM. The NAM version found in LOWTRAN6 has been updated since new scientific data has become available after its introduction in 1983. These include the following developments:

- A much more accurate parameterization of the wind dependence of large size aerosol, based on a new set of measurements.[12]
- The development of an improved multispecie aerosol growth formulation.[13]
- The inclusion of different chemical composition of the individual populations of marine aerosols. This affects both the optical properties of the aerosol and their growth properties.[3]
- An improved parameterization technique which will eliminate the necessity of knowing the air mass parameter.[14]

3.2 NOVAM'S INPUT AND OUTPUT

NOVAM has a comprehensive default system coupled with a method of estimating the "goodness" of the prediction. The philosophy behind this idea is that the model ought to be usable by everyone, even if the required input data is incomplete. However, the statistical reliability of the output should decrease as the quality of the input decreases, since that requires best estimates from other models with their inherent accuracy. This is reflected in a quality factor.

The inputs requested by NOVAM include the set of surface observations listed in Table 1 and the general profile for temperature and humidity as observed with radiosondes.

The product of NOVAM is primarily a file of the extinction and absorption coefficients at various levels in the marine atmosphere. In addition, an optional log file is produced for the user which allows an insight into what

Table 1. Surface observation data file

position	meteorological data
1	sea surface temperature (C)
2	air temperature (C)
3	relative humidity (%)
4	optical visibility (km)
5	local wind speed (m/s)
6	averaged wind speed (24 hr) (m/s)
7	air mass parameter [1..10]
8	cloud cover (tenths)
9	cloud type [1..10][3]
10	surface infrared extinction at 10.6 μm (1/km)
11	present weather in standard code [0..99]
12	height of lowest clouds (m)
13	zonal/seasonal category [1..6]

"decision" steps were taken by the model.

3.3 MODEL ARCHITECTURE

The model is based on the physical processes affecting the production, mixing, deposition and size of the aerosol within the marine atmosphere. Individual groups of aerosol with similar origin are represented by separate lognormal size distributions. All the processes which we assume to be acting on a certain group are considered to have similar effects on all particles in that group. The net optical effect produced by the aerosol is the result of the superposition of all the groups.

3.3.1. SELECTING THE PROFILE

To determine the aerosol size distribution at any particular level, one of a set of mixing profile models is used. The selection process is evident from the flow chart in Figure 2. The selected model depends on the input data available, the meteorological conditions, and the wavelength at which calculations are to be made.

Several of the modular processes in Figure 2 have yet to be formulated. These include the stable boundary layer model, a deep convection model, and a high wind stratus model. The possibility for their future existence has been planned for however in the selection process. These cases are now routed to the default mode of calculation. The modular processes which are now supported include a weak convection model, a simple mixed layer model, a sub-stratus model and a default model.

3.3.2 EXTINCTION CALCULATIONS WITH SELECTED PROFILE MODEL

In all but the sub-stratus model, the physical processes acting on the aerosol are accounted for at each level to determine the net aerosol size distribution at a nominal 80% relative humidity. The actual relative humidity at each level is determined either directly from the radiosonde data or from a default relative humidity profile generator.[15] At this point the Mie theory of light scattering and absorption from a population of aerosol is used to calculate the optical properties of the atmosphere with pre-calculated Mie integrals of extinction and absorption for the requested wavelengths and the appropriate relative humidity. This is achieved by associating these with the derived aerosol size distribution at each level. For the case of the sub-stratus model a simplified Mie calculation for each height in question is undertaken in a more specialized way.

3.4 EXAMPLES OF THE RESULTS AND COMPARISON WITH EXPERIMENTAL DATA

Figure 3 shows examples of how NOVAM estimates a profile of extinction at wavelengths in both the visible and IR bands. The meteorological profile data used as input for NOVAM was obtained form a tethered balloon platform on which also a nephelometer as well as a PMS (Knollenberg) particle spectrometer were located. Extinction at visible wavelengths is obtained directly from the nephelometer, whereas extinction at different wavelengths may be calculated, using Mie theory, from the aerosol size distributions obtained from the PMS system. The tethered balloon measurements presented in Figure 3 were taken on the upwind side of San Nicolas Island, California, on July 18, 1987.

Figure 3a is the extinction profile for wavelengths in the visible and contains extinction data measured directly with the nephelometer, extinction data calculated from the measured aerosol size distribution, and the extinction data estimated by NOVAM using the measured profile of air temperature and relative humidity. The data shows that there is a considerable amount of scatter in the measured extinction at the various altitudes. The NOVAM prediction at the visible wavelength is within the envelope of the scatter better than 75% of the time for this particular case.

Figure 3b shows the comparison between the calculations of the average extinction in the band between 3 and 5 μm using the Mie code on measured aerosol size distributions and the NOVAM prediction for 3-5 μm. Because there was no direct measure of extinction in these IR bands, only the Mie calculations of the measured size distribution are shown. A similar comparison for the 8-12 μm band and the NOVAM prediction at 10.6 μm is shown in figure 3c. The present version of NOVAM is underestimating the apparent extinction in the IR bands in the region above the inversion. This is a result of larger particles from the sea surface being mixed into the atmosphere above the apparent inversion by the process of entrainment.

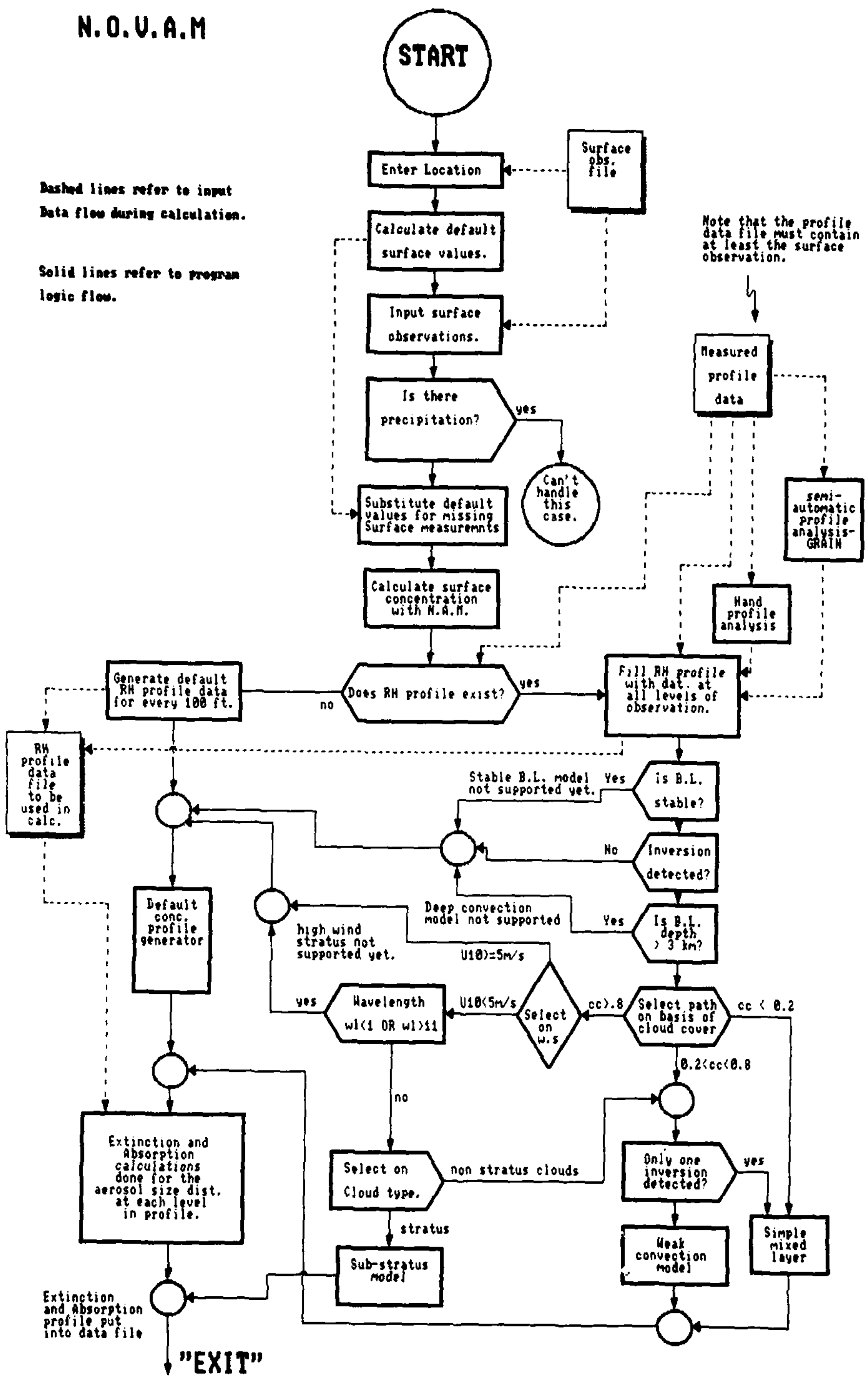


Figure 2. Flow chart for NOVAM 2.0.

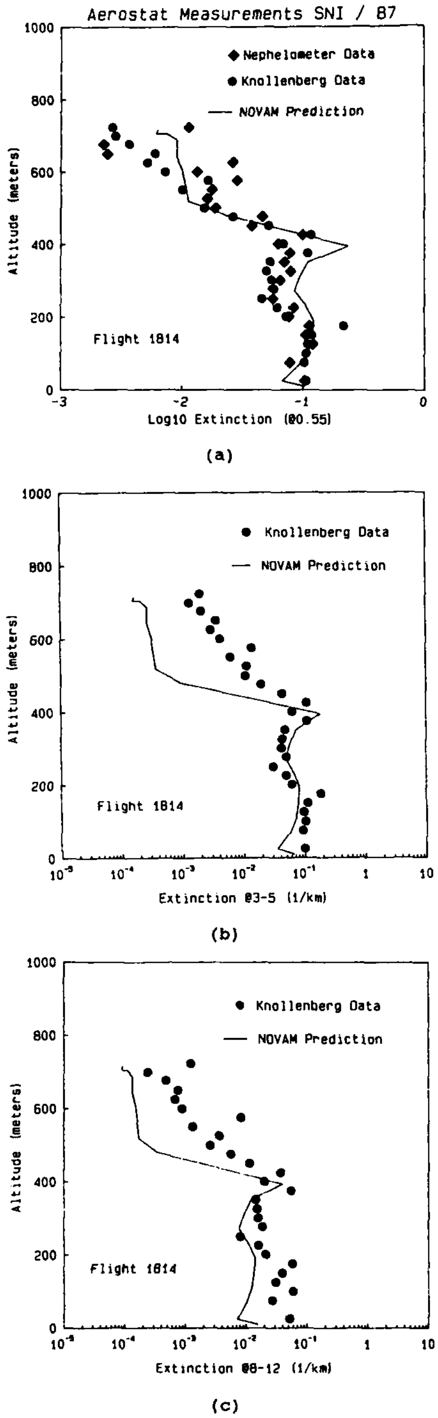


Figure 3. Comparison between extinction profiles predicted by NOVAM (solid lines) and experimental extinction profiles derived from nephelometer data (filled diamonds) or from particle size distributions using Mie theory (filled circles). The measurements were made from a tethered balloon.
(a) Comparison of the NOVAM prediction for a wavelength of 0.55 μm with a Mie-calculated profile at 0.55 μm and the extinction profile derived from the nephelometer.
(b) Comparison of the NOVAM prediction for 3.5 μm with the average Mie extinction in the 3-5 μm band.
(c) Comparison of the extinction profile as predicted by NOVAM for 10.6 μm with the Mie calculated average extinction profile in the 8-12 μm band.

4. EVALUATION AND FUTURE DEVELOPMENTS

4.1 APPROACH

The NOVAM approach as presented above, is a mixture of models developed at the author's Institutes.[1,7,10,16] The individual codes were developed for specific situations, which sets limitations to the applicability of the model. Extensions to other locations and other meteorological conditions are now major goals.

For evaluation, experimental data on the vertical structure of aerosol concentrations and optical properties are available from various experiments. Among these are aircraft aerosol measurements over the North Atlantic and the East Pacific and lidar profiles of backscatter and extinction coefficients measured over the North Atlantic and the North Sea. The geographic spreading and the variations in meteorological conditions guarantee a severe test on the usage of the model. The comparison of the calculated and observed profiles is expected to show both the strength and the weakness of the model. Improvements will be made accordingly. Some of the problems that are currently being worked on are discussed in the following sections.

4.2 RADIOSONDE SOUNDINGS

The availability of meteorological profiles is a major requirement for application of the complete model. Temperature and humidity profiles are needed to determine the height of the inversion capped mixed-layer, the temperature and humidity gradients from the surface to above the boundary layer (see Figure 1), and the cloud base. This information can be obtained from good-quality radiosonde soundings. The interpretation of the soundings to obtain the input parameters is not always straightforward. Errors in the interpretation may result in NOVAM selecting a non-representative

extinction profile model. To assist the user with the analysis of the radiosonde data, an automatic computer code is under development.

In cases when radiosonde data are not readily available, default humidity and temperature profiles are generated from the surface observation data.[15] Because these results cannot always be as good as an actual measurement the reliability of the calculated extinction profile decreases. This is expressed in the quality factor. In particular some profile parameters in Eq. (1), e.g. the entrainment rate (W_e), cannot be evaluated reliably from the default model.

4.3 RELATIVE HUMIDITY EFFECTS ON AEROSOL MIXING

In a well-mixed layer the profiles of scalar quantities can be described on the basis of surface fluxes and entrainment parameters. This does not apply to aerosol mass, because it is not a conserved scalar quantity since the size of the particles changes in response to changes in relative humidity. The freshly produced surface droplets evaporate until they are in a dynamic equilibrium with ambient humidity. This process will predominantly take place in the surface layer. In the mixed-layer the size of the aerosol particles changes because the relative humidity varies with height.

In NOVAM the particles are mixed throughout the boundary layer for a given size at 80% relative humidity (section 2.3). This is too simplified because the concentration gradients, cf. Eq. (1), also change as the particle size varies with relative humidity. At least two effects should be considered.[16] The first effect is that the effective fall velocity V_d in Eq. (1) is affected through both the change in the Stokes fall velocity and the change in the turbulent deposition velocity. The Stokes fall velocity (V_S), e.g., increases by a factor 3-4 when humidity increases from 80% to 98%. For a particle with diameter D and density ρ, V_S is given by:

$$V_S = \frac{\rho\, D^2\, g}{18\, \eta} \qquad (2)$$

where g is the gravitational acceleration and η is the dynamic viscosity. Eq. (2) shows that V_d varies with D^2, and with the particle density ρ. The particle density ρ changes with relative humidity, S, according to:

$$\rho = \rho_w + (\rho_d - \rho_w)\, g(S)^{-3} \qquad (3)$$

where ρ_w and ρ_d are the densities of pure water and of dry particles, respectively, and g(S) is the humidity correction factor that relates a particle with size D_{80}, at 80% relative humidity, to its size D at the actual ambient relative humidity[17]:

$$D = D_{80}\, g(S) \qquad (4)$$

The second effect is the shift in the particle size distribution due to humidity effects. The shift in particle size is equal for all particles of the same NOVAM mode, while different growth factors apply to different modes. However, since the mixing varies with particle size, the shape of the size distribution should change in the vertical as well. These two effects are presently being formulated for NOVAM.

4.4 AEROSOL SIZE DISTRIBUTION MODEL

The aerosol size distribution used in NOVAM is a combination of lognormal distributions describing the individual components, similar to the one used in NAM.[1] In the last decade an appreciable number of other data on the marine aerosol has become available. These were used for the new formulation of NAM that is now used in NOVAM, as described in section 3.1.

The largest particle mode in NOVAM has a mean radius of 2 μm. The aerosol extinction in both the 3-5 μm and the 8-12 μm transmission windows are predominantly determined by this 2 μm mode. The transport properties of the 2 μm particles are quite different from those of the 10 μm particles which in fact determine primarily the IR extinction properties in the 8-12 μm transmission window. This is presently not taken into account in NOVAM and the profiles for wavelengths in both IR windows have similar shapes. To describe the extinction profiles in the 8-12 μm window, it might be desirable to add another mode with a mean radius of about 10 μm.

Data on these large particles are available from surface layer measurements of aerosol size distribution profiles for particles larger than 5 μm during the HEXOS experiments,[18] covering in a wide range of wind and stability conditions. A parameterization of these particle size distributions will be attempted to take the influence of larger particles properly into account in NOVAM.

Additional improvements of the aerosol particle size distributions might be obtained from the inclusion of parameters other than mean and local wind speed, relative humidity and the air mass parameter. Monahan[19] has shown that whitecap coverage, which determines production, depends on atmospheric surface layer stability, water temperature and fetch, as well as wind speed. Further the wave properties should be considered. Wave breaking in a developing wave field is significantly different from wave breaking in an 'aged' wave field. In coastal regions the water depth and the fetch will influence the wave field.

The above considerations are important in the assessment of the present status of NOVAM. The inclusion of parameters such as fetch, stability, sea water temperature and 'wind duration' requires a new analysis of the available data. This is a major effort. On the other hand it might lead to a better parameterization of local influences and improve the applicability of NAM.

5. CONCLUDING COMMENTS

NOVAM is designed to provide realistic height variations of marine atmospheric boundary layer aerosol on the basis of dynamical and thermodynamical models for the region. There is no question, viewed from presented criticisms, that a formulation status still exists for NOVAM and that the model architecture has missing components. For all future changes data on vertical aerosol profile with complete meteorological information is needed. In spite of this current formulation status, we believe NOVAM already has merit for providing vertical extinction profiles for many geographical and meteorological regimes.

NOVAM is a candidate for forecast purposes because rate equations describe the physical processes which determine the equilibrium boundary layer. A forecast is important because mean boundary layer processes and properties, which are included in NOVAM, are continually evolving on time scales of hours.

6. REFERENCES

1. S.G. Gathman, "Optical properties of the marine aerosol as predicted by the Navy aerosol model," Opt. Eng. 22 (1983) pp. 57-62.
2. W.C. Wells, G. Gall and M.W. Munn, "Aerosol distributions in maritime air and predicted scattering coefficients in the infrared," Appl. Opt. 16 (1977) pp. 654-659.
3. S.G. Gathman, Naval Research Laboratory, "A preliminary description of NOVAM, the Navy Oceanic Vertical Aerosol Model," NRL report 9200, Washington D.C.
4. C.W. Fairall and S.E. Larsen, "Dry deposition, surface production and dynamics of aerosols in the marine boundary layer," Atmospheric Environment 18 (1984) pp. 69-77.
5. H. Tennekes and G.M. Driedonks, "Basic entrainment equations for the atmospheric boundary layer," Bound.-Layer Meteor. 20 (1981) pp. 515-531.
6. C.W. Fairall and K.L. Davidson, "Dynamics and modeling of aerosols in the marine atmospheric boundary layer," in: E.C. Monahan and G. Mac Niocaill, eds., Oceanic Whitecaps, Dordrecht, D. Reidel (1986) pp. 195-208.
7. K.L. Davidson and C.W. Fairall, "Optical properties of the marine atmospheric boundary layer: aerosol profiles," in: Ocean Optics VIII, Proc. SPIE, Vol. 637 (1986) pp. 18-24.
8. J.C. Wyngaard and R.A. Brost, "Top-down and bottom-up diffusion in the convective boundary layer," J. Atmos. Sci. 41 (1984) pp. 102-122.
9. B.A. Albrecht, "A model of the thermodynamic structure of the trade-wind boundary layer," J. Atmos. Sci., 36 (1979) pp. 90-98.
10. V.R. Noonkester, "Profiles of optical extinction coefficients calculated from droplet spectra observed in marine stratus cloud layers" J. Atmos. Sci. 42 (1985) pp. 1161-1171.
11. J.W. Fitzgerald, "Approximation formulas to calculate infrared extinction by an aerosol having a Junge size distribution," J. Appl. Meteor. 18 (1979) pp. 931-939.
12. G. de Leeuw, "Vertical profiles of giant particles close above the sea surface," Tellus 38B (1986) pp. 51-61.
13. H.E. Gerber, Naval Research Laboratory, "Relative-humidity parameterization of the Navy aerosol model (NAM)," (1987) NRL report 8956, Washington D.C.
14. A.J. Beaulieu and S.G. Gathman, To be published.
15. S.G. Gathman, Naval Research Laboratory, "Model for estimating meteorological profiles from shipboard observations," (1978) NRL report 8279, Washington D.C.
16. G. de Leeuw, "Modeling of extinction and backscatter profiles in the marine-mixed layer," Appl. Opt. 28 (1989) pp. 1356-1359.
17. J.W. Fitzgerald, "Approximation formulas for the equilibrium size of an aerosol particle as a function of its dry size and composition and the ambient relative humidity," J. Appl. Meteorol. 14 (1975) pp. 1044-1049.
18. G. de Leeuw, "Profiling of aerosol concentrations, particle size distributions and relative humidity in the atmospheric surface layer over the North Sea," Submitted to Tellus (1989).
19. E.C. Monahan "The ocean as a source for atmospheric particles," in: P. Buat-Menard, ed., The role of air-sea exchange in geochemical cycling, Dordrecht, D. Reidel (1986) pp. 129-163.

7. ACKNOWLEDGEMENTS

We wish to acknowledge other people who provided significant contributions to NOVAM. These include Dr. Juergen Richter who manages the applied research program in which it is developed, which included the observations and the physical modeling, and Drs. Chris Fairall and Hermann Gerber who were involved in numerous discussions on the direction of the formulation. GdL participated in this program while he held a National Research Council Research Associateship at the Naval Postgraduate School.

Low Level Range Coverage Performance Prediction for VHF Radar

H. Kuschel

Forschungsinstitut für Hochfrequenzphysik der FGAN e.V.
Neuenahrer Strasse 20, D-5307 Wachtberg-Werthhoven, W-Germany

Summary

At VHF radar frequencies the range coverage is not strictly limited by the quasi-optical horizon like at microwave radar frequencies but is extended due to diffraction propagation. This effect, here called beyond-the-horizon (BTH) detection capability is strongly dependent on the propagation path and thus on the terrain structure. The availability of digital terrain maps gives way to the use of computerised methods for the prediction of radar range coverage in real environment. In combination with wave propagation models suitable for diffraction at terrain structures, digital terrain data can even be used for the predication of BTH target detectability at VHF radar. Here the digital landmass system (DLMS) terrain database was used in combination with a multiple-knife-edge diffraction model to predict the diffraction attenuation between the radar and the potential target positions, especially beyond the optical horizon. The DLMS database consists of topographic height data as well as on cultural data describing the built up areas and forested areas according to their height and extension. The propagation paths extracted from the database are modelled as a sequence of diffraction screens suited for the application of a Fresnel-Kirchhoff algorithm yielding the knife-edge-diffraction attenuation. This terrain related propagation model has been verified by a large number of measurements at different frequencies.

Implemented in a fast computer system, this prediction model can be used for mission planing of air operations. Considering hostile VHF radar coverage and terrain conditon for flight path optimisation or, on the other hand it can assist in siting mobile radars for gap filling according to the actual threat situation.

Calculations of the diffraction propagation using the prediction model, described above, yield range coverage patterns in real terrain situations, allowing to quantify the BTH detection advantage of VHF radar compared to microwave radar. An experimental large wavelength radar LARA (VHF) has been built at the Forschungsinstitut für Hochfrequenzphysik to examine the potential and limitations of VHF radar for the detection of very low flying targets beyond the close horizon. Here, especially the detection of hiding helicopters by exploiting diffractive wave propagation was examined. Measurements at different VHF frequencies were carried out, to validate the results obtained by simulation.

Introduction

The ability of some classes of targets, to approach an asset undetectedly is mainly based on their low level flight capability. Using the shadowing of terrain structures to avoid detection by hostile radars is a commonly agreed tactics which is not only used by terrain following aircraft and missiles but - even to a wider extend - by helicopters which may hover behind forests or hills before attack. To counter such covered approaches, increased radar coverage has to be achieved either by additional gap filling sensors and elevated platforms, or by the exploitation of propagation effects like diffraction which occur at low radar frequencies (VHF). Since radars at these low frequencies not only offer advantages in terms of low level coverage, but do also provide preferable performance against stealth targets and ARM's. Coverage mapping for mission planing, as well as siting decisions for VHF radars should consider VHF propagation phenomena.

Wave Propagation

Refraction

The refraction of radar rays in the earth's atmosphere is caused by the variation of the refractivity in the troposphere with increasing height, resulting in a continuously bending of the radar rays towards the surface of the earth. This effect is usually taken into account using the model of a fictuous earth with increased earth radius to allow straight ray calculations. The factor, by which the earth radius is increased (multiplied) is dependent on atmospheric conditions, and thus varies for different locations and times. The factor is defined as

$$k = \frac{1}{1 + r_E \cdot dn/dh} \qquad (1)$$

where r_E is the real earth radius and dn/dh is the gradient of the refractivity with the height above surface. The factor k can range from 0.7 to 10, where 1.34 (4/3) is an often used mean value. Since the refractivity varies widely with time and space, it is difficult to apply refractivity models, other than statistical ones, however, the influence of k-factor variation is small compared to other propagation effects, especially for short range radars /1/.

Diffraction

The diffraction of electromagnetic waves at edges of obstacles is of increasing importance with increasing wavelengths. This effect was very clearly shown by Kirchhoff, using the simplified obstacle model of an ideal diffraction screen. The field strength behind the edge of an obstacle is thus described by equation 2. /2/.

$$E = \frac{E_0}{\sqrt{2}} \cdot e^{-j\frac{\pi}{4}} \left[\frac{e^{j\frac{\pi}{4}}}{\sqrt{2}} \pm (C(v) + j\,S(v)) \right] \qquad (2)$$

where E_0 is the free space field strength and

$$C(v) = \int_0^v \cos\left(\frac{\pi}{2} t^2\right) dt \qquad (3)$$

and

$$S(v) = \int_0^v \sin\left(\frac{\pi}{2} t^2\right) dt \qquad (4)$$

are the well-known Fresnel integrals. The argument of the Fresnel integrals, v, represents the physical and geometrical properties of the propagation path, i.e. the diffraction geometry referred to the wavelength.

H is the height by which the obstacle exceeds the straight line between the radar source and the selected point, where the fieldstrength is calculated, and d1 and d2 are the distances from the obstacle to the source or the selected point, respectively. λ is the wavelength of the radar emission. In this case the height H is negative. Commonly used approximations for large values of $|v|$, $|v| >> 0.5$, which is valid for locations far beyond the horizon and for very small values of $|v|$ i.e. close to the shadow boundary, $|v| < 1.5$, are given by the following equations.

$$E/E_0 = 0.225/|v| \qquad \text{for } |v| >> 0.5 \qquad (5)$$

$$E/E_0 = (1+v)/2 \qquad \text{for } |v| < 1.5 \qquad (6)$$

The figures 1 and 2 show curves of constant E/E_0 ratio versus distance behind the obstacle for a given path geometry and two different wavelengths. A comparison of the two figures clearly shows the increase of field strength behind the obstacle with increasing wavelength. However, this result is valid for the screen model which is an extremly idealised one, it can be applied to real obstacles as forest edges, hills or mountains in a first approximation. For rounded obstacles, the fieldstrength of the diffracted field decreases, compared to the ideal one, but, since the curvature which then has to be taken into account, is frequency dependent, in the case of large wavelengths, the so called knife-edge model can even be applied to rather coarse terrain structures /2/. At the FHP, a wave propagation model, based on the above described theory, has been combined with a digital terrain data base to predict the radar range performance at different frequencies in different types of terrain dependent on radar system parameters and target assumptions /3/. In figures 3 and 4 the range coverage patterns for a fictive radar site are shown at two different frequencies one in the microwave region and one at VHF, against a low flying target. The coverage patterns illustrate the low frequency detection range advantage.

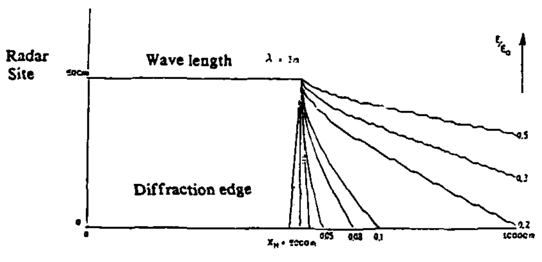


Fig. 1: Fieldstrength behind an obstacle at $\lambda = 3$ m

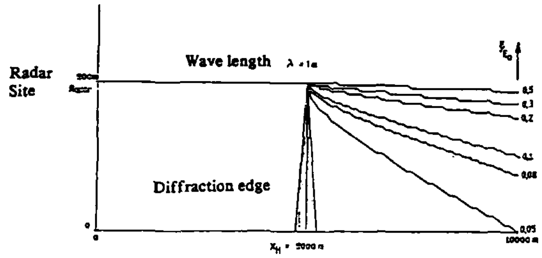


Fig. 2: Fieldstrength behind an obstacle at λ = 1 m

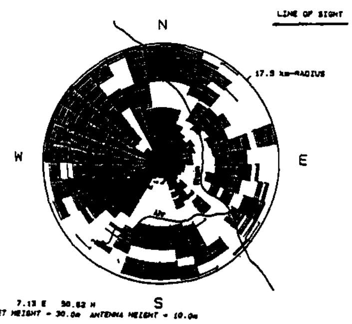


Fig. 3: Range coverage pattern line-of-sigth $\hat{=}$ frequency > 1 GHz

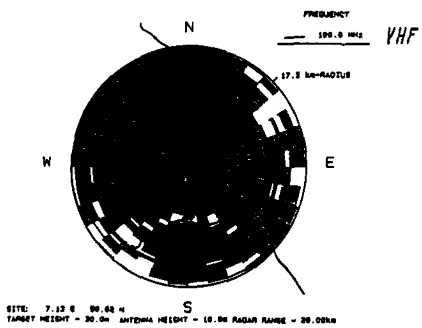


Fig. 4: Range coverage pattern at 100 MHz

Multipath Propagation

In addition to refraction and diffraction, reflections of radar rays at the surface of the earth, causing so called multipath propagation, affect radar performance. In this case, multipath means the superposition of a direct and a reflected ray at the target location which can either cause an increase by a factor of two if the rays are in phase, or a total elimation of the signal for a phase difference of 180 degrees and ideal reflection. Between these two extrema there is a wide range of resulting signal levels depending on the phase difference and the reflection coefficient. For radar application, not only the signal strength is important, but also the phase information is often needed and may be destroyed by multipath propagation. Two types of ground reflections are usually discriminated. Specular reflections which appear at smooth surfaces and diffuse reflections from rough surfaces.

The so called Rayleigh criterion

$$\lambda = 16 \cdot h \cdot \sin\phi \qquad (7)$$

gives a measure for the frequency dependent surface roughness, for which specular and diffuse reflections are approximately equal. h is the surface roughness and ϕ is the grazing angle. Equation 7 clearly indicates that for larger wavelengths specular reflection can occur for greater surface roughness or larger grazing angles, respectively. Hence, for larger wavelengths, multipath propagation is more likely to occur than for smaller ones. The contribution of the reflected ray to the interference field, however, is not only dependent on surface roughness and grazing angle, but also on the electromagnetic properties of of the surface, permittivity κ and conductivity E_r, and thus again on the frequency. For large values of $\underline{\epsilon}_r$, i.e. $|\epsilon_r|>>1$ the reflection coefficient is

$$\underline{R} = \frac{\sin\phi - \sqrt{\underline{C}}}{\sin\phi + \sqrt{\underline{C}}} \qquad (8)$$

where $\underline{C} = \frac{1}{\underline{\epsilon}_r}$

for vertical polarisation, and

$$C = \underline{\epsilon}_r$$

for horizontal polarisation
with

$$\underline{\epsilon}_r = \epsilon_r - j60 \cdot \kappa \cdot \lambda$$

(underlined values represent complex numbers) Assuming specular reflection, multipath propagation can cause serious lobing of the antenna elevation diagram in free space. The location of minima and maxima of the lobing diagram is, like the whole problem, very geometry dependent and can only be calculated, if exact knowledge of the propagation path is provided. For very small grazing angles ,provided the Rayleigh criterion for specular reflection is fulfilled, the reflection factor can be assumed to be equal to-1 which yields for the resulting fieldstrength:

$$E/E_0 = 2\ \left|\sin\left(\frac{2\pi h_r \cdot h_t}{\lambda \cdot d}\right)\right| \qquad (9)$$

where d is the distance between radar and target beeing located at heights h_r and h_t, respectively, above the reflection plane. The positions of the maxima are calculated from the condition:

$$\frac{2\pi \cdot h_t \cdot h_r}{\lambda \cdot d} = (2n-1) \cdot \frac{\pi}{2} \quad n \in N \qquad (10)$$

and those of the minima from:

$$\frac{2\pi h_t \cdot h_r}{\lambda \cdot d} = n \cdot \pi \quad n \in N_0 \qquad (11)$$

which shows that there must be a multipath minimum at the horizon. Fig.5 shows the example of an elevation diagram. This first elevation coverage minimum in the horizontal plane can be a serious restriction to radars at very low frequencies, where a lot of surface areas appear to be smooth according to the Rayleigh criterion. Hence, to limit a performance degradation by multipath, the frequency must not be choosen too low, unless the antenna can be elevated sufficiently to be out of the first minimum for the target. (like in the TV nets)

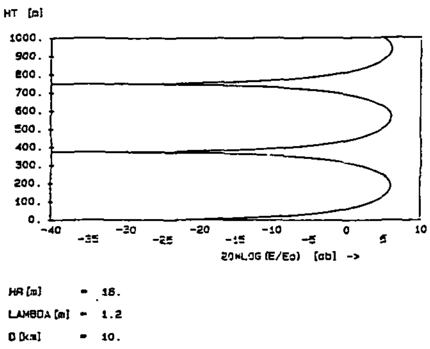


Fig. 5: Multipath loping diagram

The Experimental Radar LARA

LARA is the name of a large wavelength ecperimental radar system which was built at the Forschungsinstitut für Hochfrequenzphysik in Germany and is used for radar research in the VHF range. Verification of propagation effect modelling by measurement of height profiles is one of the main tasks, together with helicopter measurements for signature analysis and target classification. The LARA system is operated at two frequencies in the VHF range, the lower one beeing about 60 MHz and the other one at about 200 MHz in the upper VHF range.

The echo data are received, down converted to base band and recorded as I and Q-channel components for off-line processing. The radar antenna, used in the LARA system consists of a two yagi array for the 60 MHz branch and a 4 yagi array for the 200 MHz branch. Fig. 6 shows the antenna configuration.

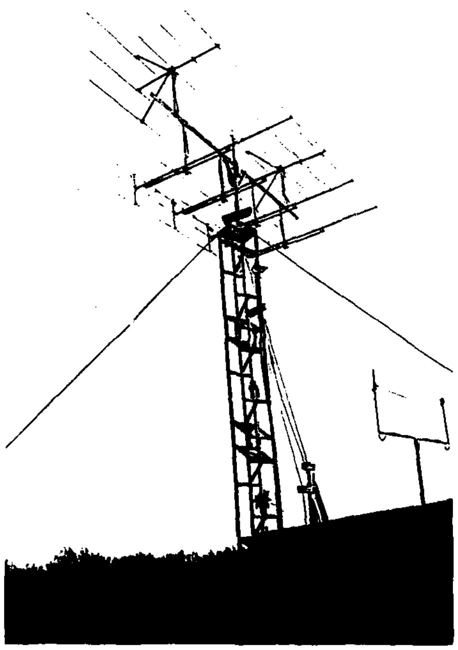


Fig. 6: LARA Antenna

Verification of propagation simulations

To quantify the beyond-the-horizon detection advantage, measured data, taken during a number of helicopter trials in Germany with the low frequency experimental radar system LARA were compared to calculated data for the same path. Figure 7 shows the geometry of the measured path where the target was monitored at positions within and beyond line-of-sight at height of -300 to+200ft with respect to the horizon line (LOS)

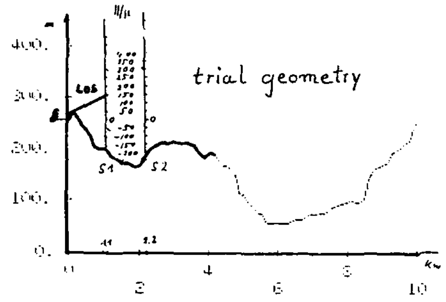


Fig. 7: Path geometry indicating target positions

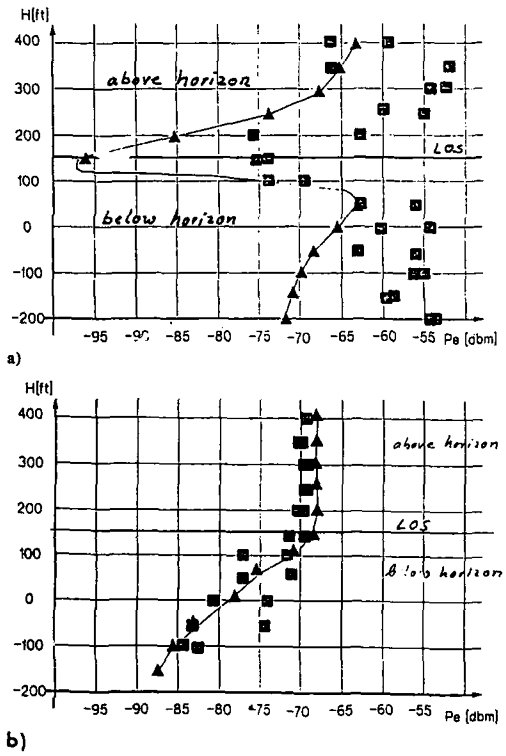
Multipath

At the heigh VHF frequencies, the surface is to rough for multipath effects, but the diffracting terrain structure can successfully be modelled as a diffraction screen. For the lower VHF-frequency, multipath propagation has to be taken into account, additionally, since the terrain surface is smooth according to the Rayleigh criterion. In contrary, the effect of beyond-the-horizon propagation is more pronounced at low frequencies. The measured height profiles of the radar echo strengt' for one path are given in fig. 8 for different VHF frequencies, where the lower VHF profile (a) shows a rather deep multipath minimum at the horizon but a regeneration zone below line-of-sight, and the higher VHF profile (b) shows no multipath but a more rapid collapse of echo strength below the horizon.

Fig. 8: a): Low VHF

b): High VHF

▲ measured
■ calculated



Physical Coverage Patterns

Combining the verified propagation model, as described above, with the terrain data base, mentioned in the beginning, range coverage plots for a given radar can be calculated for actual terrain situations.
Similar to the line-of-sight case, where a target can be detected at ranges less than the maximum free space radar range R_{max}, for locations beyond the radio horizon, a condition for detectability can be defined as follows:

Let $R_{max\ o}$ be the maximum free space radar range for given parameters like probability of detection, false alarm rate and target radar cross section, for example, then an energy surplus will be available at ranges less than $R_{max\ o}$. Hence, as long as the product of target range R and propagation factor α_D is less than or equal to the maximum free space range, the target will be detectable even beyond the horizon, according to energy calculations.

$$R \cdot \sqrt{1/\alpha_D} = R_{max\ o} \tag{12}$$

Athmospheric refraction is taken into account via the refraction index, influencing the k-faktor and thus the effective earth radaius. Since k can vary from 0.7 to about 10, the effective earth radius, for which straight ray geometrie can be applied, can vary from 4460 km to 63700 km k = 4/3 is an often used mean value. It was out of the scope of this paper, to include the examination of clutter problems.

Terrain profiles in equidistant angular spacing are extracted from the terrain database. Each path is examined for those ranges, where the condition for detectability of a target at given height is met, i.e. in equidistant range steps, the attenuation as caused by diffraction is calculated and the detectability condition is checked using equation (12).

Decision aid for VHF Radar siting

Radar coverage performance is, in general, one of the decicive features for radar siting. For microwave radars, there are serious restrictions to low level coverage due to terrain screening, so that radar sensors would have to bee sited rather close to achieve full coverage at low elevation, and would thus have some coverage redundancy at high elevations, where terrain screening is no restriction. VHF radars, however, offer some advantage against low level targets, since radar energy is diffracted at terrain structures and thus areas beyond line-of-sight are illuminated. In these areas, targets can be detected, if the reflected echo is still strong enough. In the beyond-the-horizon case, the echo power is not only a function of the target RCS, the range and the radar parameters, but also on the diffraction conditions.

In a real terrain situation, three fictitious radar sites have been chosen to achieve sufficient line-of-sight (LOS) coverage against aircraft flying at 300 ft in a terrain following mode in a given area.
This configuration is adequate for microwave radars which, in general, demand line-of-sight to the target for detection.
Figure 9 gives the combined coverage plot for the microwave sensors in the chosen positions. It is obvious that in this area three radars are necessary to achieve low level coverage, while for higher elevations, i.e. full free space range is achieved, only two sensors would be sufficient.

For the siting of the alternative low-frequency VHF-sensors, a number of possible sites were examined according to their range coverage and the optimum combination was chosen. Figure 10 which shows the VHF radar coverage, indicates that two sensors, properly sited can cover almost the whole observation area. It is important to note that quite a large percentage of the observed area is covered by diffraction which can only be taken into account using the combined terrain-propagation model for coverage calculation.

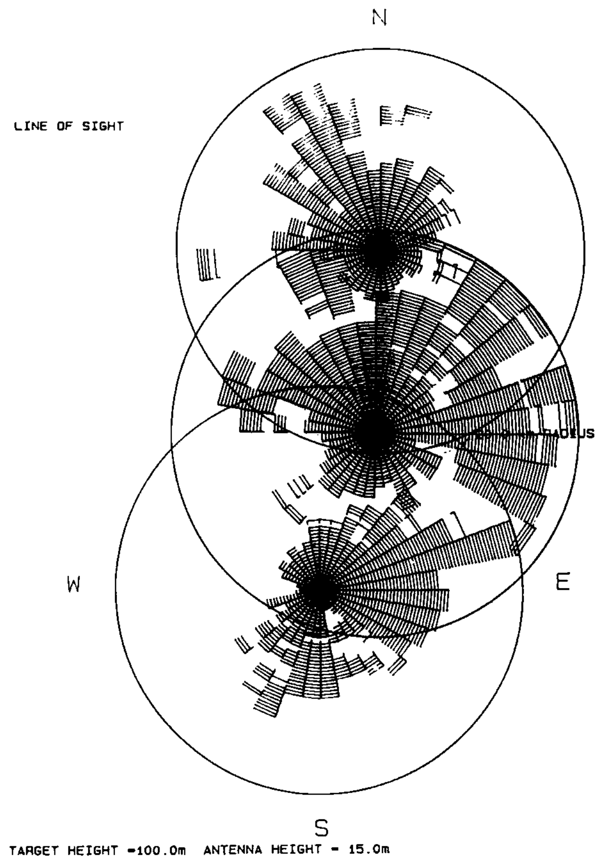


Figure 9: LOS coverage

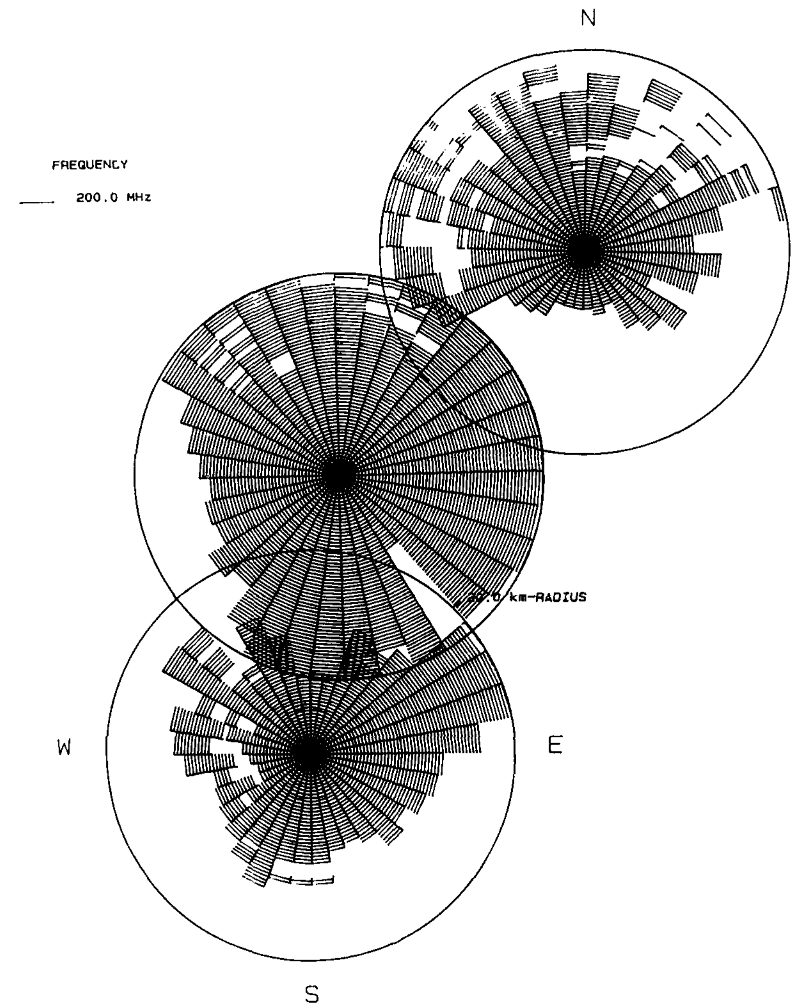


Figure 10 : VHF coverage

Flight path planing

Contrary to the problem of propet siting of radar sensors to achieve optimum areal coverage, appears the problem of detection avoidance by hostile radars. Especially radars, operating at low radar frequencies have rather good low level coverage performance, as shown before. Hence, it is important to know, which regions inside the coverage area of a netted group of radars may offer advantageous conditions for undetected approach.

Here, the most important parameter for flight path planing is the choice of flight height. Calculations of the coverage area of a fictive group of VHF-radars for different flight conditions of the approaching aircraft can be compared to decide which path and which flight height to choose.

Comparing fig.11, showing the coverage against a target flying at 60 m terrain following, to fig.12 which shows the coverage performance of the same group of VHF-radars against a target flying at 30 m above ground, indicates that in addition to flight path A at 60 m height, at the lower flight level of 30 m, another possible path for approch occurs at corridor B.

For the decision, which of these paths to use, the examination of a 3-D plot of the terrain can provide the necessary information. The 3-D-plot shown in fig. 13 has been derived from the digital terrain database, too and can even be amended by surface structure information. like forested areas, built up areas, power lines etc. During military exercises, these decisions aids have been successfully used for flight path selection for transport helicopter groups performing their task.

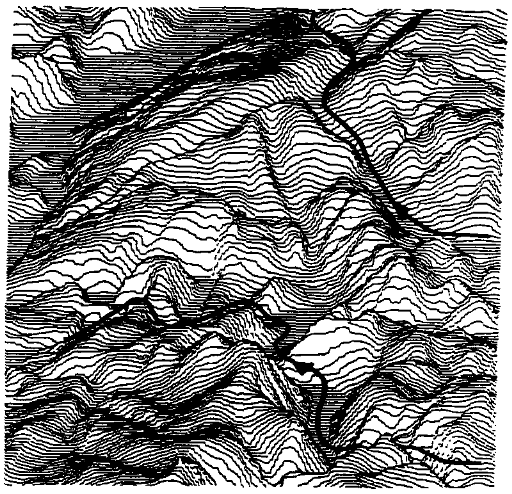
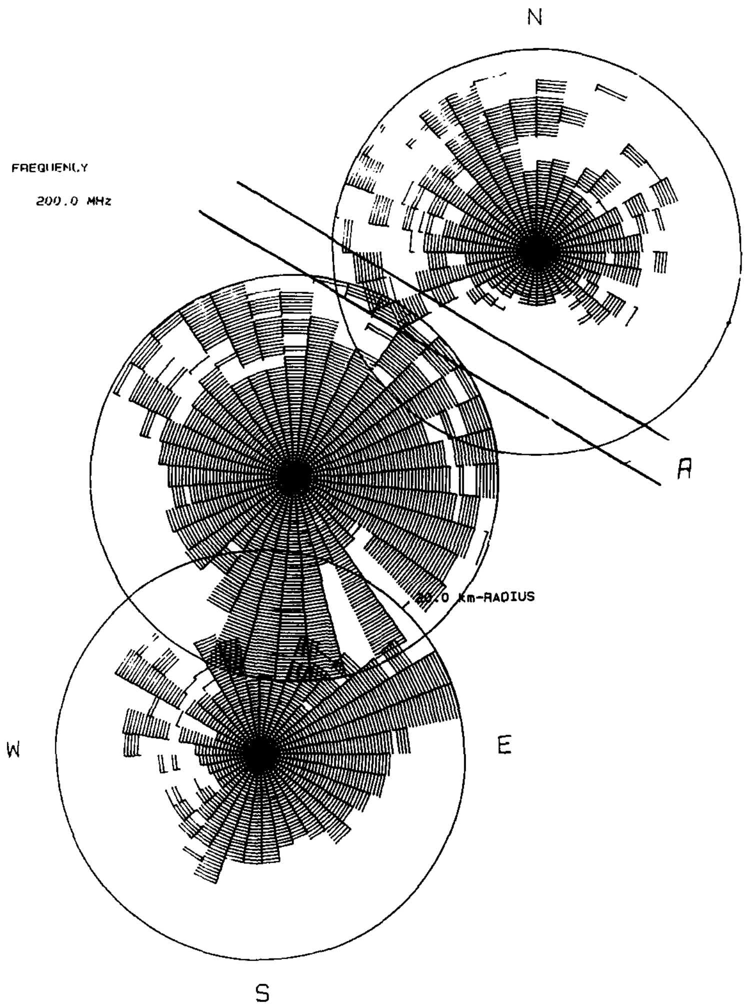


Fig. 13: 3-D terrain plot



TARGET HEIGHT = 60.0m ANTENNA HEIGHT = 15.0m RADAR RANGE = 40.00km

Figure 11 : Flight path at 60 m (200ft)

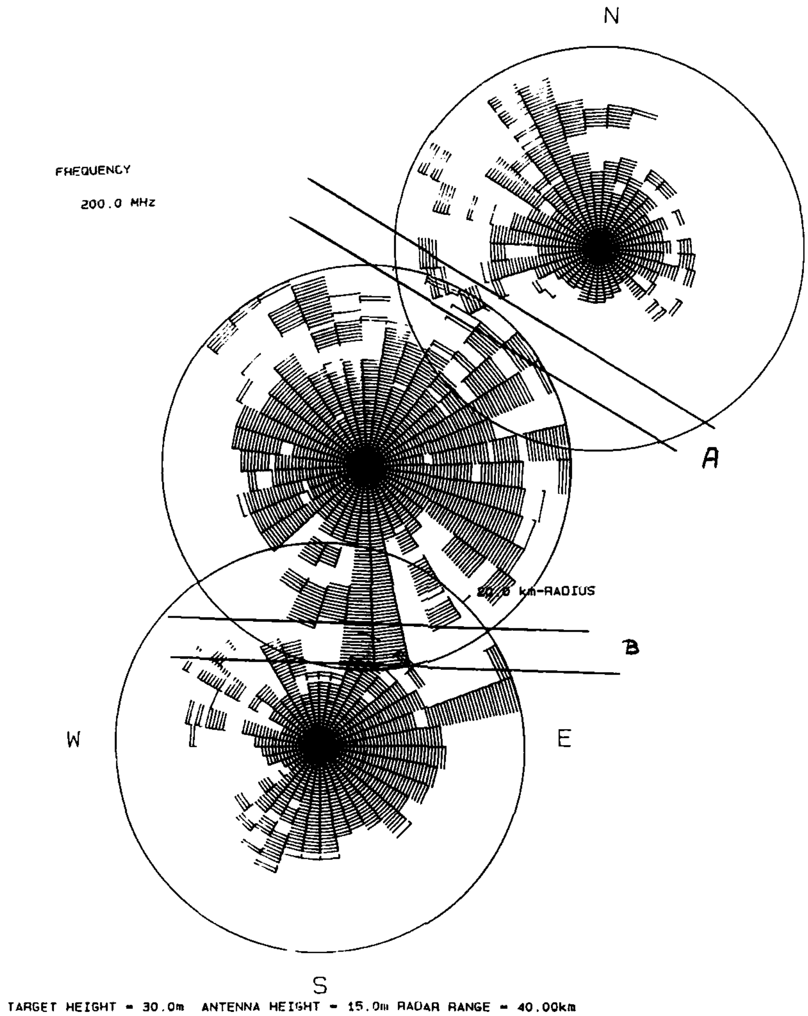


Figure 12 : Flight path at 30 m (100ft)

Conclusion

The combination of 3-dimensional terrain information with a computer supported radar coverage prediction model which includes wave propagation phenomena, has been successfully used formission planing of low level air operations. The threat of low frequency VHF radars can be taken into account as well as microwave sensors in their actual position.

On the other hand, the coverage prediction model can be uses for siting optimisation of radars and choice of preferable frequency and mode of operation. Especially for the siting of VHF/UHF radar sensors, which provide some beyond the horizon detection capability by diffractive wave propagation, siting with respect to anomalous propagation effects can obtain some additional low level coverage.

References

/1/ H. Kuschel
Mono- and Bistatic Radar Coverage Diagrams Including Diffractive Wave Propagation,
AGARD Conf. Proc. No. 407, 1987

/2/ J. Grosskopf
Wellenausbreitung I u. II,
Mannheim, Bibl. Inst. Mannheim, 1970

/3/ H. Kuschel
Radarentdeckungsreichweiten unter Berücksichtigung physikalisch-optischer Wellenausbreitung,
Forschungsinstitut für Hochfrequenzphysik,
Forschungsbericht Nr. 13-85, 1985

DISCUSSION

B. AUDONE, IT

What is the radar accuracy and resolution in the target detection?

How is it possible to take into account possible interference effects due to electrical/electronic equipment working in the vicinity of the radar site?

AUTHOR'S REPLY,

The angular resolution is dependent on the antenna aperture and is thusl different for different radar requirements in terms of mobility. The accuracy of angular measurement of target position, however, can be increased by monopulse processing. The antenna configuration that we have chosen offers the aperture for monopulse processing.

There are two approaches to counter interferences. One is to apply null steering in a phased array antenna, laying a null of the antenna receiving pattern in the direction of the interference. The other possibility is to apply spectral shaping of the radar signal, so that no radar energy is transmitted in these spectral areas, where there are interferences. The signal is then received with a matched filter, which would be a sort of notch filter. This processing would provide EM compatibility and could thus enable VHF-radar operation parallel to communication services.

F. CHRISTOPHE, FR

Do the coverage simulations you presented rely on signal to clutter and noise ratio calculations, or do you assume a perfect clutter rejection?

AUTHOR'S REPLY

We did not include clutter areas in the prediction. Our experience shows that in areas below line-of-sight there tends to be less clutter at VHF. Nevertheless, in our experimental system LARA, MTI-processing is implemented to cope with clutter.

The effect of oblate spheroidal drops on rain attenuation at 94 GHz: Comparison between theory and experiment

by
W.P.M.N. Keizer
Physics and Electronics Laboratory TNO
P.O. Box 96864
2509 JG The Haque
The Netherlands

SUMMARY

Rain induced attenuations for horizontally and vertically polarized waves at 94 GHz have been measured on a 935m link and compared with theoretical predictions. During the experiment which covered a period of two years also the raindrop size distribution along the propagation path was recorded. The theoretical predictions were based on the actually measured raindrop size distributions and assumed both spherical and spheroidal shapes for the raindrops. The calculations for oblate spheroidal raindrops showed that horizontally polarized waves are more attenuated by rain than vertically polarized waves. It was also found that the Mie scattering theory for spherical raindrops underestimates the rain attenuation for both polarizations. The comparison of the measured rain attenuation results with the predictions confirmed completely the outcome of the theoretical calculations. A remarkable improvement in the accuracy of rain attenuation modelling for horizontal and vertical polarization was noted when a spheroidal drop shape was taken into account instead of the spherical shape.

PREFACE

In an earlier paper [1] it was demonstrated that the rain attenuation at 94 GHz can fairly accurately predicted if the dropsize distribution along the propagation path is known. The rain attenuation was calculated under the assumption of a spherical shape for the raindrops. It turned out that for events with low rainfall rates the measured results agreed very well with the predictions. However, for the higher rainfall rates the comparison was less good; the measured attenuation values were on the average larger than the predicted ones. A possible reason for this discrepancy is that falling raindrops in particular the larger sizes tend have an oblate spheroidal drop shape. Larger dropsizes dominate during high rainfall rates. The effect of the distortion of the drop shape is to increase attenuation which is polarization dependent. In the earlier paper the comparison was restricted to only vertical polarization, since no experimental data for horizontal polarization was available that time.
This paper provides an extension of the earlier paper. The effect of oblate spheroidal shapes on the rain attenuation for horizontal and vertical polarization at 94 Ghz is investigated both experimentally and theoretically. Experimental rain attenuation results for the two polarizations are presented and compared with the theoretical predictions for spherical and oblate spheroidal drop shapes.

1. THEORETICAL CONSIDERATIONS

Calculation of the attenuation of electromagnetic waves caused by rain will briefly summarized. The rain attenuation of electromagnetic waves depends on the drop size distribution. The theoretical relationship between the specific attenuation c (dB/km) at wavelength k and the dropsize distribution is given by

$$\gamma = 4.3433 \cdot 10^{-3} \int_0^{\infty} N(D)\, Q_T\, (\lambda, D) dD \qquad (1)$$

where Q_T (mm^2) is the extinction cross section of a raindrop with equivalent diameter D (mm) at wavelength k and N(D)dD (m^{-3}) the number of drops per unit volume with diameter between D and D+dD.
The rainfall rate R (mm/h) may be written in terms of the dropsize distribution as

$$R = 1.885 \cdot 10^{-3} \int_0^{\infty} N(D)\; v(D)\; D^3\; dD \qquad (2)$$

where v(D) (m/s) is the fall velocity of a drop with diameter D.

One of the most commonly used dropsize distribution for calculating the rain attenuation is that proposed by Laws-Parsons [2]. Laws-Parsons derived their dropsize distribution from ground-based pellet measurements and published their results in tabulated form.
Subsequently Marshall-Palmer [3] and later Joss et al. [4] proposed distributions of a negative exponential type that give N(D) directly. All these distributions are identical in expression

$$N(D) = N_o \exp(-q/D) \qquad (3)$$

$$q = \alpha R^{-0.21}$$

where N_o and Á are constants depending on the type of rain. Instead of a single dropsize distribution, as proposed by Marshall-Palmer, Joss et al. proposed three distributions applicable to three types of rain: drizzle, widespread and thunderstorm. The constants N_o and Á for evaluating the Marshall-Palmer and Joss et al. distributions are listed in Table 1.

Table 1. Constants for evaluating dropsize distributions.

Type of dropsize distribution \ Constants	N_o $m^{-3}\ mm^{-1}$	α mm^{-1}
Marshall-Palmer	8000	4.1
Joss et al. drizzle	30000	5.7
Joss et al. widespread	7000	4.1
Joss et al. thunderstorm	1400	3.0

A significant feature of raindrops is that, although small drops fall as spheres, larger drops falling at their terminal velocities will suffer distortion due to air drag and may be assumed to be oblate spheroids with near-vertical minor axis. The ratio of this minor axis a to the major axis b is linearly related to the diameter D_e of the equivolumic sphere according to

$$a/b = 1 - 0.05\, D_e \qquad (4)$$

where a, b and D_e are expressed in mm.
Using the deformed drop model given by Eqn.(4) Dissanayake [5] calculated for a number of raindrop sizes the forward scattering complex amplitudes for horizontally and vertically polarized waves at 94 GHz with a field point matching computer program. Table 2 lists the scattering complex amplitudes for oblate spheroidal raindrops, as calculated by Dissanayake.

Table 2 Scattering complex amplitudes of oblate spheroidal raindrops at 94 GHz for zero canting angle.

Drop diameter D_e * (mm)	$S_V(0)$	$S_H(0)$
0.35	$1.0792\cdot10^{-2} - 4.0254\cdot10^{-2}i$	$1.1183\cdot10^{-2} - 4.1036\cdot10^{-2}i$
0.45	$3.1578\cdot10^{-2} - 8.8366\cdot10^{-2}i$	$3.3048\cdot10^{-2} - 9.0568\cdot10^{-2}i$
0.55	$7.9793\cdot10^{-2} - 1.6172\cdot10^{-2}i$	$8.4358\cdot10^{-2} - 1.6656\cdot10^{-1}i$
0.65	$1.7463\cdot10^{-1} - 2.5222\cdot10^{-1}i$	$1.8641\cdot10^{-1} - 2.5999\cdot10^{-1}i$
0.75	$3.2802\cdot10^{-1} - 3.3457\cdot10^{-1}i$	$3.5268\cdot10^{-1} - 3.4218\cdot10^{-1}i$
0.9	$6.3133\cdot10^{-1} - 3.7528\cdot10^{-1}i$	$6.7811\cdot10^{-1} - 3.6755\cdot10^{-1}i$
1.1	$9.8899\cdot10^{-1} - 3.0897\cdot10^{-1}i$	$1.0460 - 2.7193\cdot10^{-1}i$
1.3	$1.3002 - 3.0488\cdot10^{-1}i$	$1.3717 - 2.5275\cdot10^{-1}i$
1.5	$1.7137 - 3.7518\cdot10^{-1}i$	$1.8236 - 2.9534\cdot10^{-1}i$
1.7	$2.2540 - 4.1046\cdot10^{-1}i$	$2.3989 - 2.7108\cdot10^{-1}i$
1.95	$2.9750 - 3.7767\cdot10^{-1}i$	$3.1377 - 1.6478\cdot10^{-1}i$
2.25	$3.9151 - 4.0111\cdot10^{-1}i$	$4.1398 - 1.0211\cdot10^{-1}i$
2.55	$5.0814 - 4.1969\cdot10^{-1}i$	$5.3613 + 3.8534\cdot10^{-2}i$
2.85	$6.3572 - 3.9398\cdot10^{-1}i$	$6.6897 + 2.0551\cdot10^{-1}i$
3.15	$7.8129 - 4.1653\cdot10^{-1}i$	$8.2196 + 3.8938\cdot10^{-1}i$
3.5	$9.7480 - 3.9121\cdot10^{-1}i$	$1.0203\cdot10^{1} + 6.7952\cdot10^{-1}i$
3.9	$1.2258\cdot10^{1} - 3.9434\cdot10^{-1}i$	$1.2820\cdot10^{1} + 1.0615i$
4.3	$1.5135\cdot10^{1} - 3.677\cdot10^{-1}i$	$1.5780\cdot10^{1} + 1.5191i$
4.75	$1.8840\cdot10^{1} - 3.50\cdot10^{-1}i$	$1.9586\cdot10^{1} + 2.148i$
5.25	$2.36\cdot10^{1} - 3.1\cdot10^{-1}i$	$2.44\cdot10^{1} + 2.99i$

* Equivolumic drop diameter
n = 3.3598+1.9299i (complex refractive index at 94 GHz for T = 20° C)

From the scattering amplitudes given in Table 2 the extinction cross-sections for oblate spheroidal drops at 94 GHz can be calculated using the relation

$$Q_T = \frac{\lambda^2}{\pi} \, Re(S(0)) \qquad (5)$$

where k (mm) is the wavelength and Re{S(0)} the real part of the forward scattering complex amplitude. The resultant values are given in Table 3 together with the extinction cross-sections for spherical raindrops.

Table 3. Extinction cross-section of oblate spheroidal and spherical raindrops for vertical and horizontal polarization at 94 GHz for zero canting angle.

Equivolumic drop diameter D_e (mm)	Extinction cross section Q_T (mm^2)		
	oblate spheroidal raindrops		spherical raindrops
	vertic. pol.	horiz. pol.	
0.35	$0.3496 \cdot 10^{-1}$	$0.3621 \cdot 10^{-1}$	$0.3409 \cdot 10^{-1}$
0.45	$1.0229 \cdot 10^{-1}$	$1.0705 \cdot 10^{-1}$	$0.9835 \cdot 10^{-1}$
0.55	0.2585	0.2732	0.2453
0.65	0.5657	0.6038	0.5328
0.75	$0.1062 \cdot 10^{1}$	$0.1142 \cdot 10^{1}$	$0.1002 \cdot 10^{1}$
0.9	$0.2045 \cdot 10^{1}$	$0.2197 \cdot 10^{1}$	$0.1960 \cdot 10^{1}$
1.1	$0.3203 \cdot 10^{1}$	$0.3388 \cdot 10^{1}$	$0.3138 \cdot 10^{1}$
1.3	$0.4212 \cdot 10^{1}$	$0.4443 \cdot 10^{1}$	$0.4088 \cdot 10^{1}$
1.5	$0.5551 \cdot 10^{1}$	$0.5907 \cdot 10^{1}$	$0.5239 \cdot 10^{1}$
1.7	$0.7301 \cdot 10^{1}$	$0.7770 \cdot 10^{1}$	$0.6760 \cdot 10^{1}$
1.95	$0.9636 \cdot 10^{1}$	$1.0163 \cdot 10^{1}$	$0.8893 \cdot 10^{1}$
2.25	$0.1268 \cdot 10^{2}$	$0.1341 \cdot 10^{2}$	$0.1148 \cdot 10^{2}$
2.55	$0.1646 \cdot 10^{2}$	$0.1737 \cdot 10^{2}$	$0.1448 \cdot 10^{2}$
2.85	$0.2059 \cdot 10^{2}$	$0.2164 \cdot 10^{2}$	$0.1792 \cdot 10^{2}$
3.15	$0.2531 \cdot 10^{2}$	$0.2662 \cdot 10^{2}$	$0.2151 \cdot 10^{2}$
3.5	$0.3158 \cdot 10^{2}$	$0.3304 \cdot 10^{2}$	$0.2618 \cdot 10^{2}$
3.9	$0.3971 \cdot 10^{2}$	$0.4153 \cdot 10^{2}$	$0.3209 \cdot 10^{2}$
4.3	$0.4972 \cdot 10^{2}$	$0.5111 \cdot 10^{2}$	$0.3846 \cdot 10^{2}$
4.75	$0.6103 \cdot 10^{2}$	$0.6344 \cdot 10^{2}$	$0.4640 \cdot 10^{2}$
5.25	$0.7644 \cdot 10^{2}$	$0.7904 \cdot 10^{2}$	$0.5594 \cdot 10^{2}$

n = 3.3598+1.9299i (complex refractive index at 94 GHz for T = 20° C)

From this table it can be seen that for small drops the difference in extinction cross-section between spherical and oblate spheroidal raindrops is small, but increases with increasing dropsize and becomes substantial for the largest drops. For all dropsizes the extinction cross-section of spherical raindrops is smaller than that of spheroidal drops for both horizontal and vertical polarization.
The results of Table 3 indicate that the distortion of the drops produces significantly different attenuation for vertically and horizontally polarized waves.
One consequence that large drops fall as oblate spheroids is that the attenuation for horizontally polarized waves is always greater than that for vertically polarized waves.

2. THEORETICAL CALCULATIONS

In Fig. 1 the calculated attenuation at 94 GHz as a function of rainfall is given for the various theoretical dropsize distributions and spherical en spheroidal drop shapes using the extinction cross-sections listed in Table 3. The solid-lines represent the results for spherical raindrops, the dotted and dashed lines give the attenuation for vertically and horizontally polarized waves due to oblate spheroidal drops. As can be seen the distortion of the drop shape increases the rain attenuation for both polarizations. The largest increase is noted for horizontally polarized waves.
Fig. 2 depicts the differential attenuation between horizontal and vertical polarization versus rainfall rate. The same dropsize distributions are uses as in Fig. 1. Fig. 2 reveals that the attenuation for a horizontally polarized signal always exceeds that for a vertically polarized signal.

3. EXPERIMENTAL VERIFICATION

A 935 m, 94 GHz link at the Ypenburg airfield near the Haque was used to investigate the rain attenuation predictions made for both spherical and spheroidal drops. The propagation measurements covered a period of two years and included horizontally and vertically polarized observations. The rainfall rate was measured using three rapid response rain gauges installed along the propagation path. Registration of the dropsize distribution was done with a distrometer.

Since a two channel receiver at 94 GHz was not available no simultaneous measurement of horizontally and vertically polarized propagation at this frequency could be performed. During the first year of the experiment only horizontally polarized propagation data was gathered. Vertically polarized propagation data was collected during the second year of the experiment.

Fig. 3a shows the measured rain attenuations for vertical polarization indicated by the symbol (V) against the predicted attenuations denoted by the symbol (+). The predicted attenuations were derived from the actually measured dropsize distributions and for spherical raindrops. The solid lines in the figure represent attenuation curves based on the various theoretical dropsize distributions. The experimental data points are related to 539 events with almost homogeneous rainfall along the whole 935 m path. The homogeneity of the rainshower was verified using the three rain gauges along the path.

Fig. 3b expresses in detail the difference between the measured and the calculated rain attenuation as a function of rainfall rate for the same 539 events as in Fig 3a. The dashed line indicates the linear regression curve for the 539 data points plotted in this figure. As can be seen the regression curve does not coincide with the zero dB difference attenuation line. The departure from the zero dB attenuation line increases with increasing rainfall rate. This regression curve indicates clearly that on the average the measured attenuation values are larger than the predictions based on spherical drop shapes using simultaneously observed dropsize distributions.

The comparison between theory and experiment for horizontal polarization is given in Fig. 4 in the same way as for vertical polarization in Fig. 3. Again the predicted attenuations are based on the actually measured dropsize distributions assuming spherical drop shapes. The data points in Fig. 4a are related to 446 events characterized by uniform rainfall rate along the path. The difference between measured and predicted attenuations is depicted in detail in Fig. 4b. Also for horizontal polarization the linear regression line does not coincide with the zero dB attenuation difference line as can be seen in Fig. 4b. The measured rain attenuation values are on the average higher than the predicted ones. Comparing the results in Fig. 4b with those in Fig. 3b reveals that the difference in measured and predicted attenuation based on spherical drops is larger for horizontal than for vertical polarization. This observation is in accordance with the theoretical results presented in Fig. 1 for spherical and spheroidal drops.

The comparison between experiment and theory taken oblate spheroidal drop shapes into account is shown in Fig. 5 for vertical polarization. This figure presents the results of the same 539 events as in Fig. 3 except that the predictions are based on spheroidal drop shapes. As can be seen in Fig. 5b the linear regression almost coincides completely with the zero dB difference attenuation line indicating a nearly perfect match between theory and experiment. The same perfect match between theory and experiment is found for horizontal polarization. Fig. 6 depicts for horizontal polarization the verification between experiment and predictions made for spheroidal drop shapes using actually measured dropsize distributions. Also here the same 446 events as in Fig. 4 are considered. As can be seen in Fig. 6b the linear regression curve matches nearly completely the zero dB difference attenuation line.

The observation that horizontal polarization is more attenuated by rain than vertical polarization was definitely conformed by statistical attenuation results of a propagation experiment at 94 GHz performed over a distance of 10 km near Lorient in France [6]. During this experiment the propagation for both horizontal and vertical polarization was measured simultaneously over a period of three months from November 1981 to January 1982. The simultaneous measurement for both polarizations was made with a signal that on transmission was oblique polarized at 45 deg. On reception this signal was decomposed into a horizontally and vertically polarized component by using an orthomode transducer in the receiving antenna that was connected to a dual channel receiver.
The cumulative path attenuations due to rain for both polarizations observed during three month test period are depicted in Fig. 7. These results indicate very consistently that at 94 GHz horizontal polarization is more affected by rain than vertical polarization.

4. CONCLUSIONS

It has been demonstrated both theoretically and experimentally that a not insignificant difference in rain attenuation exists at 94 GHz for horizontal and vertical polarization. The observations and predictions have revealed that horizontally polarized waves are always more affected by rain than vertically polarized waves due to the distorted spherical shape of the raindrops. It was also found that Mie scatter theory for spherical drops underestimates the rain attenuation for both polarizations. The comparison between measured rain attenuation and predictions based on the actually measured dropsize distributions provided a very good agreement for oblate spheroidal drops.

5. REFERENCES

[1] Keizer, W.P.M.N., Snieder, J. and De Haan, C.D.: "Rain Attenuation Measurements at 94 GHz: Comparison of Theory and Experiment", AGARD Conference Proceedings No. 245, "Millimeter and Submillimeter Wave Propagation and Circuits", 1978, pp. 44-1 - 44-9.

[2] Laws, J.O., and Parsons, D.A.: "The Relation of Rain-Drop Size to Intensity", Trans. Amer. Geophys. Union, Vol. 24, 1943, pp. 452-460.

[3] Marshall, J.S. and Palmer, W. Mck.: "The Distribution of Rain Drops with Size", J. Meteorol., Vol. 5, 1948, pp. 165-166.

[4] Joss, J., Thams, J.C., and Waldvogel, A.: "The Variation of Raindrop Size Distribution at Locarno", Proceedings of the International Conference on Cloud Physics", Toronto, Canada, 1968, pp. 369-373.

[5] Dissanyake, A.W.: private communication.

[6] NATO, Defence Research Group Panel 3 on Physics and Electronics, Research Study Group 8 on Propagation and Target/Background Signatures at Millimetre Wavelengths, "Millimetre-Wave Propagation. Joint Data Analysis of the Measurement Campaigns", to be published.

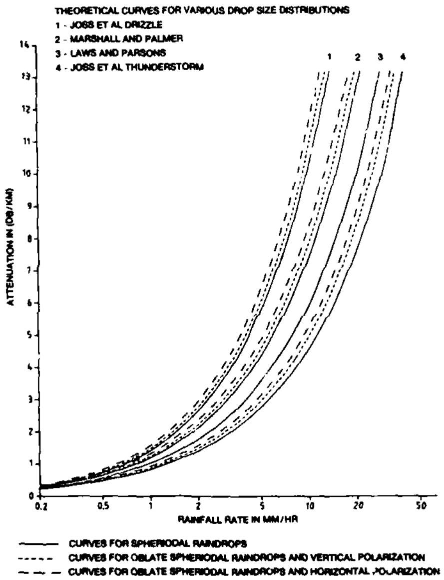


Fig. 1 Calculated attenuation for different theoretical raindrop size distributions for spherical and oblate spheriodal raindrops at 94 GHz

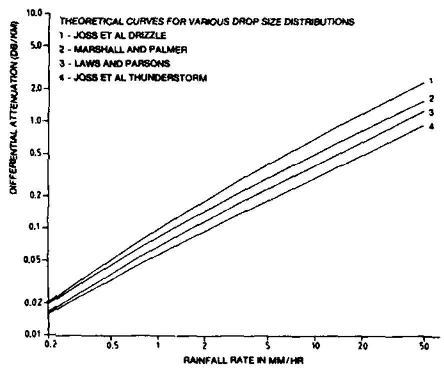


Fig. 2 Differential attenuation between horizontal and vertical polarization for different theoretical dropsize distributions as a function of rainfall rate at 94 GHz

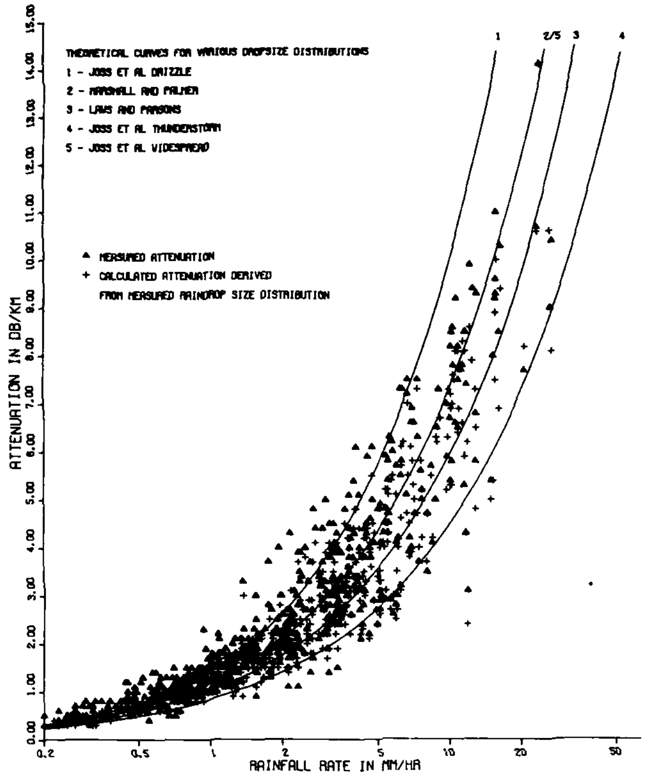


FIG. 3a MEASURED AND CALCULATED ATTENUATION VERSUS RAINFALL RATE AT 94 GHZ

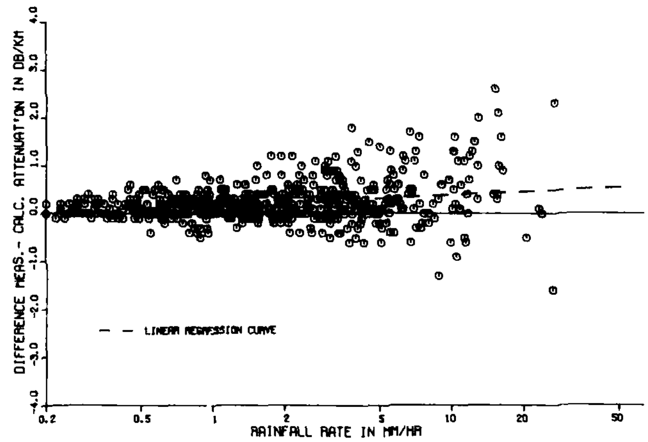


FIG. 3b DIFFERENCE BETWEEN MEASURED ATTENUATION AND CALCULATED ATTENUATION DERIVED FROM MEASURED RAINDROP SIZE DISTRIBUTION

FIG. 3 RAIN ATTENUATION AT 94 GHZ FOR VERTICAL POLARIZATION: COMPARISON OF THEORY AND MEASUREMENT FOR SPHERICAL RAINDROPS

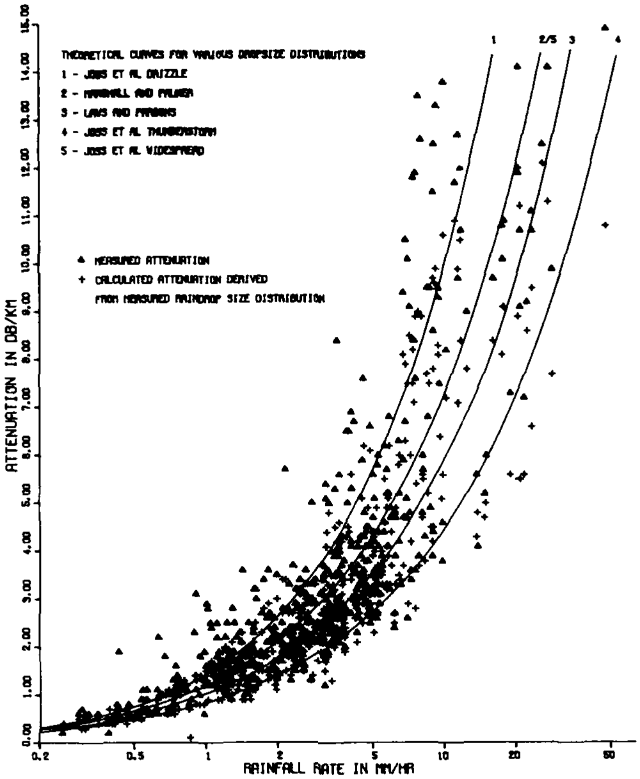


FIG. 4a MEASURED AND CALCULATED ATTENUATION VERSUS RAINFALL RATE AT 94 GHZ

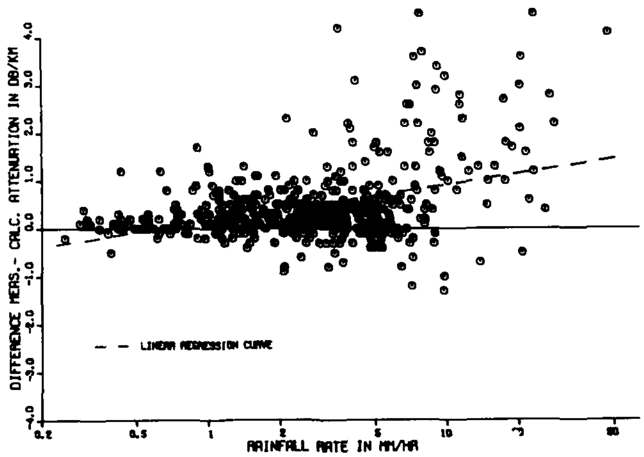


FIG. 4b DIFFERENCE BETWEEN MEASURED ATTENUATION AND CALCULATED ATTENUATION DERIVED FROM MEASURED RAINDROP SIZE DISTRIBUTION

FIG. 4 RAIN ATTENUATION AT 94 GHZ FOR HORIZONTAL POLARIZATION: COMPARISON OF THEORY AND MEASUREMENT FOR SPHERICAL RAINDROPS

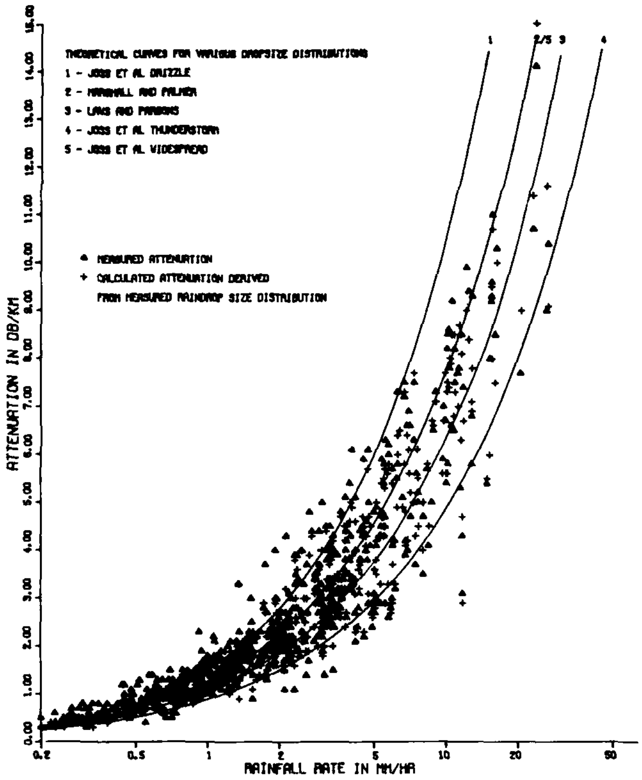


FIG. 5a MEASURED AND CALCULATED ATTENUATION VERSUS RAINFALL RATE AT 94 GHZ

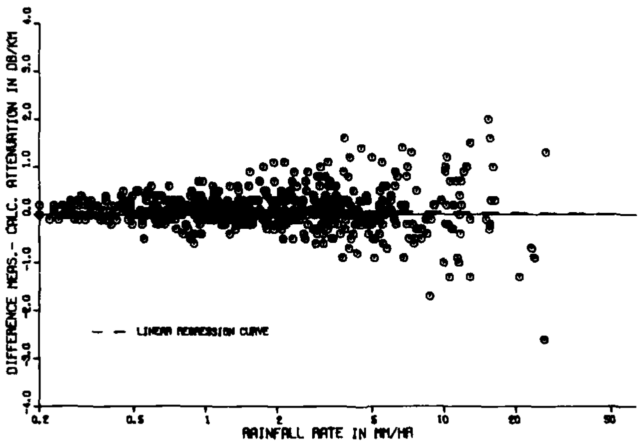


FIG. 5b DIFFERENCE BETWEEN MEASURED ATTENUATION AND CALCULATED ATTENUATION DERIVED FROM MEASURED RAINDROP SIZE DISTRIBUTION

FIG. 5 RAIN ATTENUATION AT 94 GHZ FOR VERTICAL POLARIZATION: COMPARISON OF THEORY AND MEASUREMENT FOR OBLATE SPHEROIDAL RAINDROPS

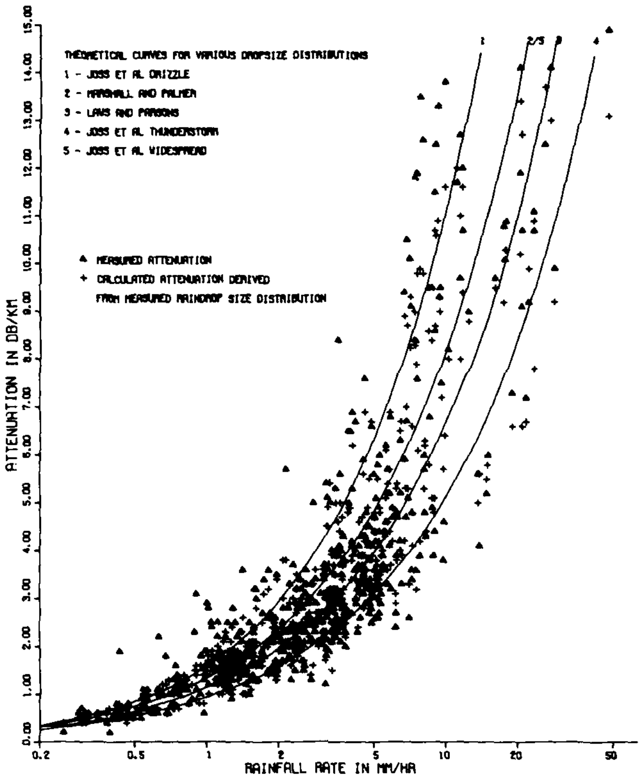


FIG. 6a MEASURED AND CALCULATED ATTENUATION VERSUS RAINFALL RATE AT 94 GHZ

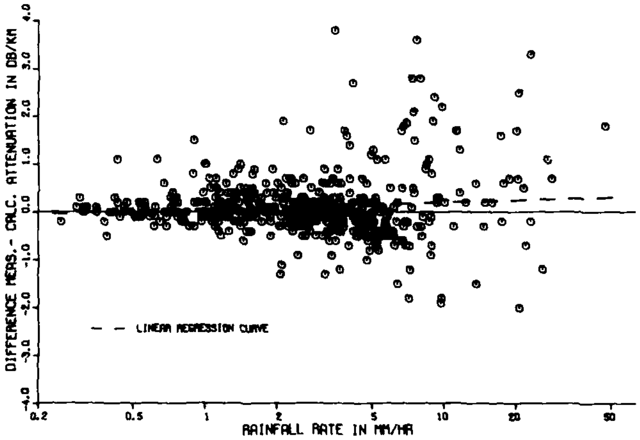


FIG. 6b DIFFERENCE BETWEEN MEASURED ATTENUATION AND CALCULATED ATTENUATION DERIVED FROM MEASURED RAINDROP SIZE DISTRIBUTION

FIG. 6 RAIN ATTENUATION AT 94 GHZ FOR HORIZONTAL POLARIZATION: COMPARISON OF THEORY AND MEASUREMENT FOR OBLATE SPHEROIDAL RAINDROPS

Rain attenuation statistics at 94 GHz for horizontal and vertical polarization along a 10 km link

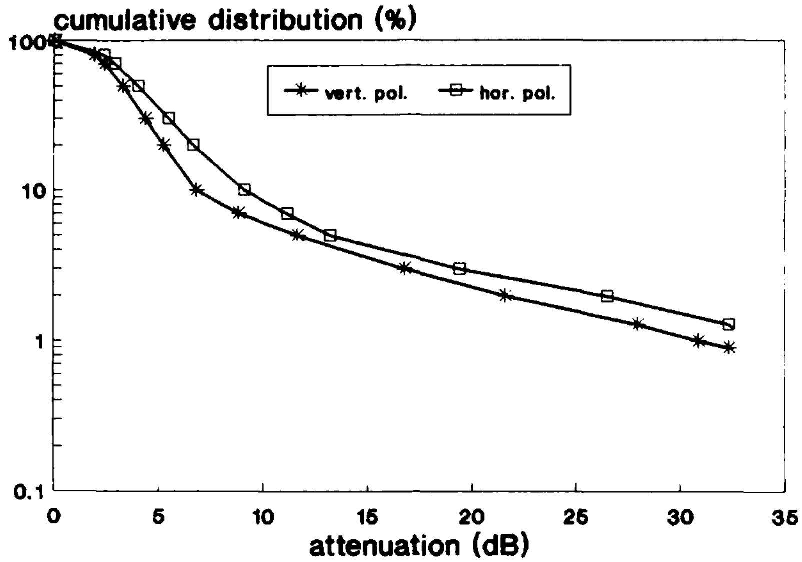


Fig. 7

DISCUSSION

D. HÖHN, GE

What mathematical methods and/or approximations were applied to calculate non-spherical rain droplet MIE extinction coefficients?

AUTHOR'S REPLY

The numerical values for the extinction cross-sections of oblate spheroidal raindrops were supplied by A. W. Wissanyake of Bradford University. He used a field point matching computer program for the computations.

T. FITZSIMONS, UK

For heavy rainfall rates, say above 40 mm/hour, the rain-cell size becomes important and may be smaller even than the 935 m link you used. Could the spread of difference between calculated and measured values at high rainfall rates be due to the rainfall rate not being constant along the path?

AUTHOR'S REPLY

All experimental attenuation values in the comparison ween theory and experiment are related to events with uniform rainfall along the whole path. Differences in the predicted and measured attenuations are probably caused by errors in the measurement of the dropsize distributions.

However, heavy rainfall rates were never measured because of the large integration time (83 sec) used for the measurement of the rainfall rates and dropsize distribution.

MITIGATING DOPPLER SHIFT EFFECT IN HF MULTITONE DATA MODEM

Yaşar SONLU
3.Hava İkmal Bakım Merkezi Komutanlığı
Fabrika Müdürlüğü
06375 Etimesğut-Ankara-TÜRKİYE

SUMMARY

Digital communications over High Frequency radio channels are getting important in recent years. Current HF requirements are for data transmission at rates 2.4 kbps or more to accomodate computer data links and digital secure voice. HF modems which have been produced to meet these speeds are two types; serial modems and parallel modems.

On the other hand, the HF sky-wave communication medium, the ionosphere, has some propagation problems such as multipath and Doppler shift. In this paper, the effect of Doppler shift in a parallel modem which employs Differential Quadrature Phase Shift Keying (DQPSK) modulation is considered and a correction method to mitigate the Doppler Shift effect is introduced.

1. INTRODUCTION

Once the limitations, capabilities, and cost of satellites became clearer, mainly in the 1970s, greater attention was again focused on HF radio as a classical communications medium for long-distance transmission links [1]. Especially, in the military communications field, the growing recognition of vulnerability of satellite links to interception and even destruction has helped stimulate the development of HF radios [2]. HF radio uses frequencies in the range from 2 to 30 megahertz. At these frequencies, communications beyond line-of-sight is due to refructive bending of the radio wave in the ionosphere from ionized layers at different elevations [3]. This refraction phenomena in the ionosphere results in the reflection of the radio waves from the ionosphere. Thus, very long range communications can be achieved by using low-power and low-cost radio equipment.

Most current combat radios still operate as analog systems. This makes them liable to intercept, regardless whether the message has been disguised, scrambled or inverted. Certain regularities in the transmission pattern are bound to occur and this facilitates an intercept. There is no other solution than to digitize all signals and use appropriate processing, not only for greater security but also because digitizing offers considerable advantages over current analog signal formats. One advantage is that partially received signals can be reconstructed. Above all, digitized signals lend themselves readily and without exception to all types of transmission, which certainly cannot be claimed for analog systems [4].

The need to transmit data over HF radio long distance communication links has grown from low speed low quality links to a requirement for both high integrity and high speed data transmission [5].

In the past forty years, there has been an uninterrupted trend toward the digitization of communications, with far reaching consequences in terms of improved reliability, increased operational speed, reduced equipment size, freedom from calibration problems, improved ability to mechanize complicated signal processing algorithms, etc [6].

Until comparatively recently, securing speech was carried out by analogue techniques which ensured effective scrambling of the voice signals whether in time or frequency or both. Inevitably a greater degree of security is inherent in new techniques and now digital systems have evolved [7]. In digital systems, speech is converted to digital signal and then encrypted and finally transmitted over a HF radio link by a suitable modem. Because the HF channel bandwidth is only about 3 kHz, this limits the modem's data rate to 2.4 kbps for most practical purposes [8]. One of the possible technique to encode speech at bit rates of 2.4 kbps and lower is the Linear Predictive Coding (LPC) technique [9].

HF modems for high speed data such as 2.4 kbps digitized voice have evolved toward multiple-tone signaling carrying differential phase shift keying. Serial data modems provide economical service over wirelines; but HF radio networks, multipath delay usually causes intersymbol interference. Consequently, parallel tones having durations much longer than the expected multipath delays are favored, even though parallel channels have been an expensive implementations. Digital data processing now minimizes the parallel channel cost penalty; however, the linearity requirements on RF power amplifiers, combined with the high peak-to-average power ratio of parallel tones, degrate system performance compared to that with constant envelope modulation modems. New exploratory serial modem developments may provide performance improvements over the parallel modem [10].

The parallel tone modems divide the data into low rate parallel sub-channels so that nonadaptive techniques can be used and intersymbol interference can be avoided.

The serial tone approaches use PSK transmissions and some form of decision feedback equalization in an adaptive receiver structure [11]. In serial tone modems, channel estimation method can also be used instead of decision feedback equalization method [12].

Ionospheric HF channels are fading channels as a result of dispersive phenomena that take place in both the time and frequency domains. Frequency dispersion is caused by Doppler phenomena, whereas, in most cases, time dispersion is caused by multipath propagation [13]. The range of values for time dispersion are about between 0.5 ms and 5 ms [14]. On the other hand, the values of frequency dispersion is generally under 4 Hz (∓2 Hz) and with strong interference, such as magnetic storm, the values up to 12 Hz have been observed [15]. Also, at high latitudes, abnormal Doppler shifts can be observed due to the variable nature of the high-latitude ionosphere [16]. Physical effects that could cause the Doppler shift include vertical motion of the propagation path associated with gravity waves or other large plasma irregularities moving laterally through the ionospheric midpath region at the height of reflection [17].

The following sections of the paper investigates the effect of Doppler shift in a typical High Speed (2.4 kbps) parallel modem which employs Differential Quadrature Phase Shift Keying modulation, and to mitigate the Doppler shift effect in these modems, a correction method is introduced.

2. HF PARALLEL MODEMS

One of the high speed parallel modem which employs DQPSK modulation is defined in the MIL-STD 188C. Its basic concept is to make the baud interval longer than the maximum multipath distortion (about 8 ms). Indeed, the parallel tone modem has become an established way of lessening the effect of intersymbol interference caused by multipath for high data rates. In operation, the transmitted data is split into 16 Differentially Phase Shift Keyed tones and one additional pilot tone for frequency offset correction due to radio mistuning or Doppler shift. By changing the phase angle of each tone simultaneously at a rate of 75 times a second, a data throughput of 2.4 kbps can be reached while keeping a rate of only 75 baud [18]. The message signal contains tones which each carry 2 bits of data. At every frame boundary, the tone phases are shifted by a multiple of 90 degrees with respect to the previous frame to represent the data. Sometimes, this kind of modulation is called as "Time Differential Quadrature Phase Shift Keying" [19].

Let us consider and examine that kind of modem. The 17 tones are spaced equally in a 3 kHz HF Single Side Band (SSB) channel, and the pilot tone which is used for frequency offset correction is normally at the middle of the channel as shown in Figure 1. In

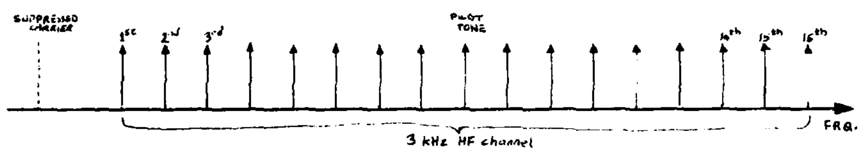


Figure 1. Arrangement of the tones in a 3 kHz HF/SSB channel.

modulation process, the serial bit sequence is taken 32 bits at a time and transmitted in pairs (2 bits) with each tone, so that with a frame rate of 75 baud, 2.4 kbps speed is accomplished. In sky-wave communication, the HF/SSB signal radiated from a transmitter travels upwards from the Earth's surface, and reflected by the ionosphere, and then arrives at the receiver antenna. Depending upon the distance between the transmitter and the receiver, arrival of the HF/SSB signal takes a certain time. This time delay creates phase errors in the received signal. We can assume that time delays for all tones in the same frame are equal to one another. However, the same amount of time delay makes different phase error for each tone in the same frame because each tone has a different frequency. To explain this point more clearly, let us give an example. Suppose the frequency of the first tone is 5,000,000 Hz, and that of the second tone is 5,000,155 Hz, and the time delay is 4 ms. The phase errors created by the 4 ms time delay are

$$\text{Phase error in the first tone} \triangleq C_1 = \int_0^{t_d} 2\pi f_1 \, dt = 2\pi f_1 t \Big|_0^{t_d}$$

$$= 2\pi (5{,}000{,}000)\ 0.04$$

$$= 2\pi\ 200{,}000$$

$$= 0 \text{ rad}$$

$$= 0^\circ$$

Phase error in the second tone $\triangleq C_2 = \int_0^{t_d} 2\pi f_2 \, dt = 2\pi f_2 t \Big|_0^{t_d}$

$$
\begin{aligned}
&= 2\pi\,(5{,}000{,}155)\ 0.04 \\
&= 2\pi\ 200{,}006.2 \\
&= 2\pi\,(\underbrace{200{,}006}_{0} + 0.2) \\
&= 2\pi\ 0.2 \\
&\cong 1.2566 \text{ rad} \\
&\cong 72^\circ
\end{aligned}
$$

As shown above, the phase errors due to the time delay are different from each other. If we consider two successive frames radiated from the transmitter, we can represent the phases for each tone as follows:

First Frame (transmitted at t_1)	Second Frame (tx.at $t_2 = t_1 + 13$ ms)
1^{st} tone $\longrightarrow \Phi_T 1(t_1)$	1^{st} tone $\longrightarrow \Phi_T 1(t_2)$
2^{nd} tone $\longrightarrow \Phi_T 2(t_1)$	2^{nd} tone $\longrightarrow \Phi_T 2(t_2)$
. .	. .
. .	. .
. .	. .
15^{th} tone $\longrightarrow \Phi_T 15(t_1)$	15^{th} tone $\longrightarrow \Phi_T 15(t_2)$
16^{th} tone $\longrightarrow \Phi_T 16(t_1)$	16^{th} tone $\longrightarrow \Phi_T 16(t_2)$

After a certain time delay, t_d, when these two frames arrive at the receiver, we can represent the tone phases of the received signal in a similar way as follows:

First Frame (received at t_1+t_d)	Second Frame (received at t_2+t_d)
1^{st} tone $\longrightarrow \Phi_R 1(t_1+t_d)$	1^{st} tone $\longrightarrow \Phi_R 1(t_2+t_d)$
2^{nd} tone $\longrightarrow \Phi_R 2(t_1+t_d)$	2^{nd} tone $\longrightarrow \Phi_R 2(t_2+t_d)$
. .	. .
. .	. .
. .	. .
15^{th} tone $\longrightarrow \Phi_R 15(t_1+t_d)$	15^{th} tone $\longrightarrow \Phi_R 15(t_2+t_d)$
16^{th} tone $\longrightarrow \Phi_R 16(t_1+t_d)$	16^{th} tone $\longrightarrow \Phi_R 16(t_2+t_d)$

or, written as

First Frame (received at t_1+t_d)	Second Frame (received at t_2+t_d)
$\Phi_R 1(t_1+t_d) = \Phi_T t(t_1) + C_1$	$\Phi_R 1(t_2+t_d) = \Phi_T 1(t_2) + C_1$
$\Phi_R 2(t_1+t_d) = \Phi_T 2(t_1) + C_2$	$\Phi_R 2(t_2+t_d) = \Phi_T 2(t_2) + C_2$
. .	. .
. .	. .
. .	. .
$\Phi_R 15(t_1+t_d) = \Phi_T 15(t_1) + C_{15}$	$\Phi_R 15(t_2+t_d) = \Phi_T 15(t_2) + C_{15}$
$\Phi_R 16(t_1+t_d) = \Phi_T 16(t_1) + C_{16}$	$\Phi_R 16(t_2+t_d) = \Phi_T 16(t_2) + C_{16}$

To obtain the information bits from the received signal, the phase value of each tone in

the first frame is subtracted from the phase value of each respective tone in the second frame so that we have

$$\Phi_R 1(t_2+t_d) - \Phi_R 1(t_1+t_d) = \Phi_T 1(t_2) + C_1 - \Phi_T 1(t_1) - C_1 = \underbrace{\Delta\Phi\, 1(t_2,t_1)}_{\text{2 bit data}}$$

$$\Phi_R 2(t_2+t_d) - \Phi_R 2(t_1+t_d) = \Phi_T 2(t_2) + C_2 - \Phi_T 2(t_1) - C_2 = \underbrace{\Delta\Phi\, 2(t_2,t_1)}_{\text{2 bit data}}$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\Phi_R 15(t_2+t_d) - \Phi_R 15(t_1+t_d) = \Phi_T 15(t_2) + C_{15} - \Phi_T 15(t_1) - C_{15} = \underbrace{\Delta\Phi\, 15(t_2,t_1)}_{\text{2 bit data}}$$

$$\Phi_R 16(t_2+t_d) - \Phi_R 16(t_1+t_d) = \Phi_T 16(t_2) + C_{16} - \Phi_T 16(t_1) - C_{16} = \underbrace{\Delta\Phi\, 16(t_2,t_1)}_{\text{2 bit data}}$$

Total information obtained from the arrival of each frame is 32 bits.

Here we shold state that the modulation process in the transmitter and the demodulation process in the receiver are actually applied to the baseband signals, not to the RF signals. In the transmitter, modulated baseband signal is up converted to the RF channel frequency, and in the receiver the demodulation process takes place after down converting the received RF signal to a baseband signal. In the foregoing analysis about the phase errors created by time delays and how to get the information bits from the received RF signal was referred to the RF signals instead of referring to baseband signals. However, it can be easily justified that the phase relations among the tones of a HF/SSB signal and that of the respective baseband signal are similar, so that we can either refer to RF signals or respective baseband signals when making explanation about the subject. Another point that should be stated here is that the propagation time delay for the second frame has been taken as equal to that for the first frame. This is a reasonable assumption because for the same channel frequency and for a short duration, the propagation path length is almost unchanged.

Now let us return to our analysis again. The pilot tone does not contain any information; it is an unmodulated tone. In the receiver, the frequency of the received pilot tone is measured, and if a deflection from the nominal value is detected then the required correction is made. In the foregoing analysis we assumed that there was no frequency deflection in the received signal. However, in reality, the frequency deflection coming either from Doppler shift phenomenon in the ionosphere or from the radio equipment themselves may not be perfectly corrected due to several reasons. These are

a) Frequency deflection correction is made in certain steps. Firstly, the frequency of the pilot tone is measured; secondly, the measured value is compared with the nominal value and if a deflection is detected, then the oscillator frequency in the receiver is changed accordingly. Each step stated here takes some definite time, so that in the case of rapid Doppler shift variations in the ionosphere, the receiver may not keep up with the frequency fluctuations.

b) In any physical system, there are always limitations in accuracy and resolution. These limitations in receivers create frequency error.

c) Fading caused by time and/or frequency dispersion may effect the pilot tone so that the frequency correction can not be made correctly for the duration of fading.

Hence, in the receiver a frequency deflection may exist. Frequency deflection creates additional phase errors in the received signal. These phase errors can be calculated from the formula

$$\text{Phase error due to frequency deflection} \triangleq K = \int_{t_1}^{t_n} 2\pi\Delta f(t)\,dt$$

where t_1 = starting time of the message
t_n = the time at which the phase error is calculated
$\Delta f(t)$= frequency deflection

Examination of the formula reveals that the amount of phase error is depend on both the elapsed time from the beginning of the message, and the frequency deflection. The frequency deflections for all tones in the same frame can be taken as equal to one another, so that the created phase errors for all tones in that particular frame will be equal to one another. When calculating the phase error for each frame of the received message, if we take the beginning time of each frame as time t_n, then we have

$$\text{Phase error for the first frame} \triangleq K_1 = \int_{t_1}^{t_1} 2\pi \Delta f(t)dt = 0$$

$$\text{Phase error for the second frame} \triangleq K_2 = \int_{t_1}^{t_2} 2\pi \Delta f(t)dt = \int_{0}^{13ms} 2\pi \Delta f(t)dt$$

$$\text{Phase error for the third frame} \triangleq K_3 = \int_{t_1}^{t_3} 2\pi \Delta f(t)dt = \int_{0}^{26ms} 2\pi \Delta f(t)dt$$

For example, if we assume a fixed frequency deflection of 2 Hz, the phase errors for the second and third frame will be

$$K_2 = \int_0^{13ms} 2\pi\, 2dt = 4\pi t \Big|_0^{13ms} = 4\pi(0.013) = 0.163 \text{ rad} = 9.36^\circ$$

$$K_3 = \int_0^{26ms} 2\pi\, 2dt = 4\pi t \Big|_0^{26ms} = 4\pi(0.026) = 0.326 \text{ rad} = 18.72^\circ$$

As shown above, the phase errors due to the frequency deflection for different frames are not the same.

Taking into account both types of phase errors due to propagation time delay and frequency deflection, the phases of the first three frames of the received signal can be given

First Frame	Second Frame	Third Frame
$\Phi_r 1(t_1) + C_1$	$\Phi_r 1(t_2) + C_1 + K_2$	$\Phi_r 1(t_3) + C_1 + K_3$
$\Phi_r 2(t_1) + C_2$	$\Phi_r 2(t_2) + C_2 + K_2$	$\Phi_r 2(t_3) + C_2 + K_3$
.	.	.
.	.	.
.	.	.
$\Phi_r 15(t_1)+ C_{15}$	$\Phi_r 15(t_2)+ C_{15}+ K_2$	$\Phi_r 15(t_3)+ C_{15}+ K_3$
$\Phi_r 16(t_1)+ C_{16}$	$\Phi_r 16(t_2)+ C_{16}+ K_2$	$\Phi_r 16(t_3)+ C_{16}+ K_3$

One can see that, after applying the DQPSK demodulation process to the received signal, the demodulated signal includes phase errors due to frequency deflection. Therefore, DQPSK technique is incapable of eliminating the phase errors due to frequency deflection in the receiver. In the next section of the paper, a correction method to mitigate the frequency deflection effect will be introduced.

3. CORRECTION METHOD

As explained in the previous section, DQPSK technique can not eliminate the phase

errors due to frequency deflection in the receiver. These phase errors can only be eliminated with a vertical subtraction between adjacent tones of the same frame. In the transmitter, if the information is coded both in the phase difference between two adjacent tones of the same frame and in the phase difference between the same tones of two successive frames, then both types of phase errors in the received signal can be eliminated by inverse decoding process. This is the basis of the correction method. To explain the correction method more clearly, let us consider the first and second frame of the transmitted message. We define "$\delta\Phi$" as the phase difference between any two adjacent tones.

First Frame of Tx. Signal		Second Frame of Tx. Signal	
$\Phi_T 1(t_1)$		$\Phi_T 1(t_2)$	
	$> \delta\Phi_T 1(t_1)$		$> \delta\Phi_T 1(t_2)$
$\Phi_T 2(t_1)$		$\Phi_T 2(t_2)$	
	$> \delta\Phi_T 2(t_1)$		$> \delta\Phi_T 2(t_2)$
$\Phi_T 3(t_1)$		$\Phi_T 3(t_2)$	
$\vdots$		$\vdots$	
$\Phi_T 15(t_1)$		$\Phi_T 15(t_2)$	
	$> \delta\Phi_T 15(t_1)$		$> \delta\Phi_T 15(t_2)$
$\Phi_T 16(t_1)$		$\Phi_T 16(t_2)$	
	$> \delta\Phi_T 16(t_1)$		$> \delta\Phi_T 16(t_2)$
$\Phi\ 17(t_1)$		$\Phi_T 17(t_2)$	

where $\delta\Phi_T n(t) = \Phi_T n{+}1(t) - \Phi_T n(t)$

$\Phi_T n(t)$ is the phase of the n^{th} tone transmitted at time t

In the correction method, the information is coded in the phase difference between $\delta\Phi_T$s instead of Φ_Ts as in DQPSK technique. To be able to send 32 bits of information per frame, it is necessary to use 17 tones as shown above. When these two transmitted frames are received with two types of phase errors by the receiver, the first step to get the information bits from the received signal is to do vertical subtraction, that is, to take the phase difference between adjacent tones in the same frame as follows:

Vertical Subtraction for the First Frame

$$\Phi_R 2(t_1+t_d)-\Phi_R 1(t_1+t_d)=\Phi_T 2(t_1)+C_2-\Phi_T 1(t_1)-C_1=\delta\Phi_T 1(t_1)+C_2-C_1 \triangleq \delta\Phi_R 1(t_1+t_d)$$

$$\Phi_R 3(t_1+t_d)-\Phi_R 2(t_1+t_d)=\Phi_T 3(t_1)+C_3-\Phi_T 2(t_1)-C_2=\delta\Phi_T 2(t_1)+C_3-C_2 \triangleq \delta\Phi_R 1(t_1+t_d)$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\Phi_R 16(t_1+t_d)-\Phi_R 15(t_1+t_d)=\Phi_T 16(t_1)+C_{16}-\Phi_T 15(t_1)-C_{15}=\delta\Phi_T 15(t_1)+C_{16}-C_{15} \triangleq \delta\Phi_R 15(t_1+t_d)$$

$$\Phi_R 17(t_1+t_d)-\Phi_R 16(t_1+t_d)=\Phi_T 17(t_1)+C_{17}-\Phi_T 16(t_1)-C_{16}=\delta\Phi_T 16(t_1)+C_{17}-C_{16} \triangleq \delta\Phi_R 16(t_1+t_d)$$

Vertical Subtraction for the Second Frame

$$\Phi_R 2(t_2+t_d)-\Phi_R 1(t_2+t_d)=\Phi_T 2(t_2)+C_2+K_2-\Phi_T 1(t_2)-C_1-K_2=\delta\Phi_T 1(t_2)+C_2-C_1 \triangleq \delta\Phi_R 1(t_2+t_d)$$

$$\Phi_R 3(t_2+t_d)-\Phi_R 2(t_2+t_d)=\Phi_T 3(t_2)+C_3+K_2-\Phi_T 2(t_2)-C_2-K_2=\delta\Phi_T 2(t_2)+C_3-C_2 \triangleq \delta\Phi_R 2(t_2+t_d)$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\Phi_R 16(t_2+t_d)-\Phi_R 15(t_2+t_d)=\Phi_T 16(t_2)+C_{16}+K_2-\Phi_T 15(t_2)-C_{15}-K_2=\delta\Phi_T 15(t_2)+C_{16}-C_{15} \triangleq \delta\Phi_R 15(t_2+t_d)$$

$$\Phi_R 17(t_2+t_d)-\Phi_R 16(t_2+t_d)=\Phi_T 17(t_2)+C_{17}+K_2-\Phi_T 16(t_2)-C_{16}-K_2=\delta\Phi_T 16(t_2)+C_{17}-C_{16} \triangleq \delta\Phi_R 16(t_2+t_d)$$

The next step of the demodulation process is to subtract the results of the vertical subtraction for the first frame from that for the second frame as follows:

$$\delta\Phi_R 1(t_2+t_d)-\delta\Phi_R 1(t_1+t_d)=\delta\Phi_T 1(t_2)+C_2-C_1-\delta\Phi_T 1(t_1)-C_2+C_1 = \Delta\delta\Phi_T 1(t_2,t_1)$$

$$\delta\Phi_R 2(t_2+t_d)-\delta\Phi_R 2(t_1+t_d)=\delta\Phi_T 2(t_2)+C_3-C_2-\delta\Phi_T 2(t_1)-C_3+C_2 = \Delta\delta\Phi_T 2(t_2,t_1)$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$\Phi_R 15(t_2+t_d)-\delta\Phi_R 15(t_1+t_d)=\delta\Phi_T 15(t_2)+C_{16}-C_{15}-\delta\Phi_T 15(t_1)-C_{16}+C_{15}= \Delta\delta\Phi_T 15(t_2,t_1)$$

$$\Phi_R 16(t_2+t_d)-\delta\Phi_R 16(t_1+t_d)=\delta\Phi_T 16(t_2)+C_{17}-C_{16}-\delta\Phi_T 16(t_1)-C_{17}+C_{16}= \Delta\delta\Phi_T 16(t_2,t_1)$$

Each $\Delta\delta\Phi_T$ term represents 2 bits of information, so that totally, 32 bits of information is obtained per frame.

4. CONCLUSION

HF modems which have been produced to meet 2.4 kbps or more to accomodate data links and digital secure voice are two types; serial modems and parallel modems. Serial modems use single tone and some type of equalising technique, whereas parallel modems uses multitones. In this paper, the effect of Doppler shift in a typical high speed parallel modem which employs DQPSK technique has been investigated. It has been shown that DQPSK technique can eliminate the phase errors due to propagation time delay, but it is incapable of eliminating the phase errors due to Doppler shift or frequency offset in some cases. To overcome this problem, a correction method has been introduced. In this method, the information is coded both in the phase difference between two adjacent tones in the same frame and in the phase difference between two succesive frames. In the receiver, demodulation takes place in two steps. First step is to take the phase difference between adjacent tones in the same frame; and second step is to take the phase difference between the results of the first step for two adjacent frames. With this method, both types of phase errors can be eliminated. Since serial modems use single tone, the correction method is not applicable to them.

REFERENCES

1. Lindner, J., "Technical possibilities", Comm. Inter., May.1987, pp.28
2. Lilley, J., "HF radio on the rise", Comm. Eng. Inter., Nov.1987, pp.48
3. Monsen, P., "Modern HF communications, modulation and coding", AGARD-LS-145, 1986, pp.8-1
4. Geisenheyner, S., "A survey of secure battlefield communications systems", Armada Inter., 1/1988, pp.32
5. Fahy, J., "The revival of HF data communications", Comm. Inter., Dec.1980, pp.46
6. Grossi, M.D., "An introduction to error-control coding with application to HF communications", AGARD-LS-145, 1986, pp.7-1
7. Ashby, J., "Communications security", MILTECH, 5/86, pp.192
8. Wiesner, L., "Telegraph and data transmission over shortwave radio links", Third edition, John Wiley and Sons, Siemens, 1984, pp.67
9. Blakley, J.R., J.R.Brouk, J.B.Kleijn, R.J.Myles, B.C.Prorok, and R.J.Zurawaski, "An efficient implementation of the LPC-10e speech coding algorithm", IEEE MILCOM 87, Vol.2, 1987, pp.0560
10. Wilson, Q.C., "Developments in HF equipment and systems: mobil and portable terminals", AGARD-LS-145, 1986, pp.10-1
11. Monsen, P., "Modern HF communications, modulation and coding", AGARD-LS-145, 1986, pp.8-7
12. Lilley, J., "HF radio on the rise", Comm. Eng. Inter., Nov.1987, pp.52
13. Belrose, J.S., R.Bultitude, D.Clark, R.W.Jenkins, W.Lauber, G.Nourry, N.M.Serinken, and G.Venier, "An overview of Canadian radio propagation/comminications technologies research", AGARD-CPP-442, Oct.1988, pp.2-2
14. Monsen, P., "Modern HF communications, modulation and coding". AGARD-LS-145, 1986, pp.8-2
15. Braun G., "Planning and engineering of shortwave links", Second edition, John Wiley and Sons, Siemens, 1986, pp.45
16. Darnell, M., "The design of static and mobile HF communication systems", AGARD-LS- 145, Mar.1986, pp.4-5
17. Wagner, L.S., J.A.Goldstein, and W.D.Meyers, "Wideband probing of the trans-auroral HF channel", IEEE MILCOM 87, Vol.3, 1987, pp.0877
18. Lilley, J., "HF radio on the rise", Comm. Eng. Inter., Nov.1987, pp.52
19. Buding, H.E, "A microprocessor controlled parallel modem with FDPSK", AGARD-CP-363, Oct.1984, pp.38-1

A One-Dimensional Shipboard Model for Forecasting Refractive Effects in the Planetary Boundary Layer

Paul M. Tag
Software Coordinator, TESS Project
Naval Environmental Prediction Research Facility
Monterey, CA 93943-5006
U.S.A.

SUMMARY

In order to forecast the refractivity structure in the atmosphere, the accompanying temperature and moisture structures must be forecast. As a means of providing shipboard-generated forecasts of low-level temperature and moisture structures in the planetary boundary layer, the Navy has converted a sophisticated, turbulence model to a micro-computer. This model, called the Navy Over-Water Local Atmospheric Prediction System (NOWLAPS), is part of the Tactical Environmental Support System (TESS), a shipboard, automated, environmental diagnosis/prediction system. NOWLAPS is a one-dimensional, second order closure model that can provide forecasts over water for temperature, moisture, and winds up to heights of 2 km., and out to 24 hrs. From these temperature and moisture distributions, the refractivity profile can be generated.

1. Introduction

A knowledge of the current refractivity structure in the lower atmosphere is of obvious tactical and strategic importance. Even better is a forecast of that refractivity structure. To determine that refractivity structure, either a measured or a predicted "sounding" of temperature and moisture is necessary. Measurements are usually obtained by means of a radiosonde, a weather balloon that radios back measurements to the user as it ascends through the atmosphere. Such a sounding is valid only for the column of air through which the balloon travels; however, over the ocean, away from land influences, horizontal homogeneity is not always a bad assumption. On the other hand, since most naval operations are conducted in the vicinity of a coastline, new efforts in determining wave propagation characteristics are focusing on the more complicated situation involving horizontal variations in temperature and moisture structure.

In contrast to a diagnosis of the current atmospheric temperature and moisture structure conditions, the forecast of those same conditions requires a fundamental understanding of the physics that controls the atmosphere and, in the case of low-level refractivity, the planetary boundary layer (PBL). The model that we chose for conversion for shipboard use was one that had initially been developed at the Naval Environmental Prediction Research Facility (NEPRF) as a research tool for the study of the PBL. This model is a one-dimensional (1-D), second-moment closure turbulence model, in which the closure consists of parametrizing the third moment terms and solving prognostic equations for the second moment terms (see Burk, 1977, 1980). This type of model has, in recent years, received considerable attention and application to the simulation of the PBL. For purposes of PBL simulation, our model includes moisture, as well as parametrizations for both long and shortwave radiation.

The means by which NOWLAPS could be run onboard ship, and the stimulation behind this project, was the Tactical Environmental Support System (TESS). TESS was conceived in 1982 as a minicomputer-based environmental diagnosis/forecast system for shipboard use. The term "environmental" means that TESS is more encompassing than the limited area of atmospheric diagnosis/prediction. Ocean prediction, remote sensing, and ship response, among others, are relevant to the environmental applications of TESS.

Most forecast models require huge mainframe computers to run, mostly because they are three-dimensional--covering the entire globe, or more limited areas in high resolution. In addition to these computer requirements, such models require sophisticated schemes to ingest the large amounts of data required for initialization. One such forecast model whose requirements could be met by the anticipated computer power and data availability in TESS is the 1-D, PBL model addressed above.

A logical question stemming from the above is the following: why should we be concerned with placing any numerical forecast capability on ship when central-site computers could handle more sophisticated and accurate models, and produce the results more quickly. The numerical results could then be transmitted by radio or satellite to the ship. There are two main reasons. First, a product generated with locally available data could be processed locally without using communication channels going to and from the central site. And second, more importantly, a locally-generated product would be available even during a communications outage, as could happen during wartime.

2. NOWLAPS Development

In adapting the 1-D NEPRF PBL model for shipboard use, we patterned its development after a central-site version that had been considered for operational use. This version, the Navy Operational Local Atmospheric Prediction System (NOLAPS), is run on a central-site computer at the Fleet Numerical Oceanography Center (FNOC), the Navy's forecast center. (See Burk and Thompson, 1982 for a detailed description of NOLAPS.) NOLAPS can produce a 24 hr. forecast for any oceanic location on the earth, both in the northern and Southern Hemisphere.

The primary developmental work that went into NOLAPS centered on the inherent limitation of a 1-D model--the fact that one-dimensionality implies horizontal homogeneity. In order for NOLAPS to be a successful forecast tool, a method of accounting for horizontal advection had to be devised. That method is based on using the local time rates of change as defined by the Navy Operational Global Atmospheric Prediction System (NOGAPS), the Navy's global forecast system run at FNOC:

$$\frac{d\bar{\phi}}{dt} = \frac{\bar{\phi}(t+\Delta t) - \bar{\phi}(t)}{\Delta t}$$

In this expression, $\bar{\phi}$ is a NOGAPS-predicted variable (such as temperature and moisture) at on of the global grid points. That change over a twelve-hour period (based on an initial analysis and a twelve-hour forecast) is interpreted to be the synoptic, large-scale, advective change that is then added to the appropriate NOLAPS model equation:

$$\frac{\partial\bar{\phi}}{\partial t} = \frac{\partial}{\partial z}\,(-w'\phi') + \xi\,(\frac{d\bar{\phi}}{dt})$$

where
$$\xi = \begin{cases} 1 & z > z_{850} \\ (z/z_{850})^2 & 0 \le z \le z_{850} \end{cases} .$$

The factor ξ is a weighting function that diminishes the computed synoptic tendency at heights above 850 mb. This reduction is necessary to eliminate redundant computations in the PBL, where the NOGAPS PBL parameterization, although much less sophisticated than that in NOLAPS, is nonetheless operative.

3. Adaptation to a Desktop Computer

The adaptation of the NOLAPS model to a desktop computer environment (and thus becoming NOWLAPS) has gone through several iterations, starting in 1984. At that time, the Navy's standard computer used in the ship's meteorological spaces was the Hewlett-Packard (HP) 9845. Our benchmark goal for running NOWLAPS on this machine was 1 hour for a 24 hour forecast; that goal was achieved. Starting in 1986 with the implementation of TESS 1.0, a more powerful machine, the HP 9020, became available. The running time for that 24 hour forecast was reduced to several minutes, a more realistic time frame for operational shipboard use.

Aside from the mechanics of converting a mainframe-running NOLAPS into a desktop computer-running NOWLAPS, a method for determining the large-scale tendency terms was a primary concern. The initial version of TESS to be installed onboard ship would not have the large-scale fields from which the tendencies could be automatically computed, as can be done in NOLAPS at FNOC. Instead, a manual method was substituted; this method use values interpolated, by hand, from the temperature, moisture, and wind hardcopy, NOGAPS analyses and forecast charts.

4. Forecast Accuracy

In an attempt to simulate the use of NOWLAPS on an aircraft carrier in the open ocean, weather station ship Charlie, located at latitude 52.7N, longitude 35.5W in the north Atlantic, was used to provide the test case data. NOWLAPS, using the manual tendency method, was compared to forecasts from both NOLAPS and persistence. Persistence is a forecast that the current conditions will continue into the future. For many atmospheric forecasts, beating persistence is not a trivial accomplishment and is indicative of some skill.

We chose RMS errors, based upon the verifying ship soundings, as the basis for comparing the three forecasts. The 1-D model simulations were run to a height of 2250 m, with the NOWLAPS model having 25 grid points within that domain. Although wind speed and direction were also evaluated in the study, only the results for temperature and specific humidity are of concern here for refractivity. Root mean square (RMS) errors, based upon the verifying soundings, were computed for the entire 2250 m domain for each of the 27

cases. Overall, both NOWLAPS and NOLAPS produced mean RMS errors that are lower for temperature and moisture (2.3 C, 0.8 G/KG--NOLAPS; 3.0 C, 1.0 G/KG--NOWLAPS) than for persistence (3.3 C, 1.2 G/KG). Also, in the mean, NOLAPS is clearly superior to NOWLAPS. Because the models are ostensibly the same, the differences must lie primarily in the errors from the manual interpretation of the tendency terms.

An analysis of the RMS errors by themselves do not reveal the total picture, however. For NOWLAPS and persistence, Figures 1 and 2 show the number of cases that have RMS errors within incremental groupings. Note that, for both temperature and moisture, the standard deviation for persistence is much larger than for NOWLAPS. This result shows that the NOWLAPS results are more tightly grouped at the lower end of the RMS scale, and thus more reliable than the RMS errors by themselves would suggest.

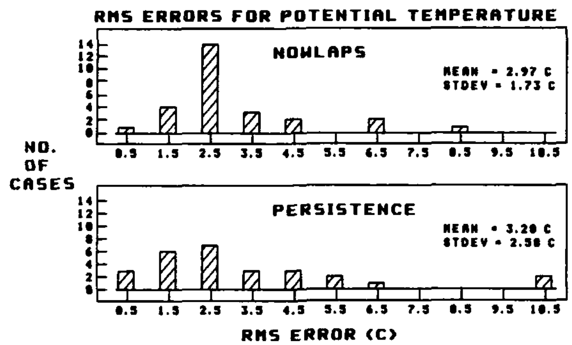


Figure 1. Number of cases versus RMS error for NOWLAPS and persistence, for potential temperature (see text).

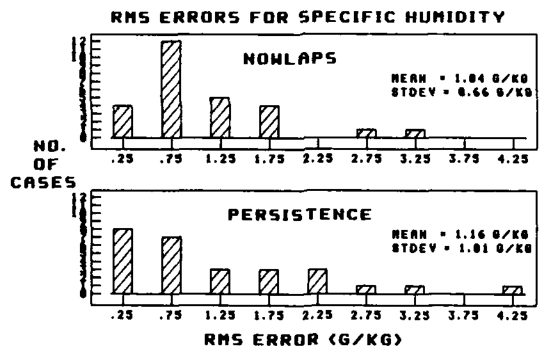


Figure 2. As in Figure 1, except for specific humidity.

5. Future Plans

The above results show that there is clear room for improvement. While beating persistence is a noteworthy accomplishment, more useful results are still to be obtained. For these reasons, we are currently developing the second generation of this model, NOWLAPS 2.0.

The current method for determining the tendency terms suffers from two weaknesses. First, as the above testing revealed, a manual substitution for the automated method produces less accurate results (not to mention the time-intensive nature of this substitution). Second, there is a fundamental flaw in our use of the time rates of change from the global model. Those time rates of change do include turbulence processes and radiation effects, albeit less sophisticated than in NOWLAPS, from the global model physics. There is an attempt to remove the boundary layer effects (see Section 2 above), but the method is ad hoc and unsatisfactory.

Two approaches are being investigated to eliminate the above difficulty. In the first, NOWLAPS would be run in a trajectory, rather than eulerian, mode. In other words, the model would be initialized upstream with a sounding representing the air that will eventually arrive at the forecast site. One advantage of this approach is that the

variable sea surface temperature along the trajectory path could be used as the lower boundary condition. Although conceptually simple, this lagrangian approach has considerable additional complexity involved in determining that upwind location. A second, perhaps more manageable solution to an improved NOWLAPS rests with a revised tendency term extracted from NOGAPS--a tendency term that includes, specifically, **only** the horizontal advection.

Both of the above alternatives have been tested in a revised version of NOLAPS. Initial testing shows that the results from the two methods are comparable, and superior to results from the original NOLAPS. The second method has obvious implementation advantages over the Lagrangian technique.

Earlier reference was made to the first version of TESS that is now operative in the fleet. What makes either of the above two NOWLAPS updates possible is the implementation of the third phase, TESS (3) (the first two phases are both based upon the HP9020). TESS(3) will be a major upgrade of the system and will include not only a new, more powerful computer, but also the transmission of central site, 3-D, environmental fields into the shipboard computing system, thus creating a complete environmental data base for the battlegroup. These, among other additional features, will permit NOWLAPS 2.0 to be a significant upgrade to the current shipboard version.

6. REFERENCES

Burk, S. D., 1977: The moist boundary layer with a higher order turbulence closure model. J. Atmos. Sci., **34**, 629-638.

Burk, S. D., 1980: Refractive index structure parameters: time-dependent calculations using a numerical boundary-layer model. J. Appl. Meteor., **19**, 562-576.

Burk, S. D., and W. T. Thompson, 1982: Operational evaluation of a turbulence closure model forecast system. Mon. Wea. Rev., **110**, 1535-1543.

DISCUSSION

J. GOODMAN, US

You mention that the degree of complexity is substantial when proceeding from the Eulerian to the Langrangian approach. Can this be avoided by exploiting actual radio data to "update" the local M profiles to the remote region for which a forecast is required?

AUTHOR'S REPLY

Jay Rosenthal at the Pacific Missile Test Center (PMTC) has conducted research relating satellite imagery to refractivity and duct heights. NOWLAPS is a meteorological model that must be initialized with meteorological data; the refractivity profile in NOWLAPS is diagnosed from that data. In addition, you would not know the location of the radio data that you would need until you had completed the upwind trajectory computations.

WEATHER SATELLITE AND COMPUTER MODELING APPROACHES TO ASSESSING PROPAGATION OVER MARINE ENVIRONMENTS

by
Jay S. Rosenthal, Roger A. Helvey, Steven W. Lyons,*
Alan D. Fox, Richard Szymber and Lee Eddington
Pacific Missile Test Center
Geophysics Division (Code 3250)
Point Mugu, California 93042-5000
USA

SUMMARY

Exploiting electromagnetic propagation in the Navy's marine environment requires an ability to assess and predict atmospheric refractive structure over the ocean. Efforts have been underway to develop such capabilities by correlating duct existence, height and intensity to synoptic and mesoscale weather features. In addition to the latter conventional parameters, meteorological satellite data, through pattern recognition has been valuable in inferring important aspects of ducting conditions.

More recently, a technique was developed to derive duct height information directly from computer-processed infrared (IR) satellite data over ocean regions capped by stratified low clouds. This approach, still in development, also provides a means of measuring the horizontal variability of duct height which is crucial to determining the appropriateness and useability of various propagation models for predicting systems performance.

Inferring the impact of horizontal variability over the ocean has been facilitated by use of range-dependent raytrace techniques which allow inputs from such diverse sources as local radiosonde observations; predicted marine layer depth (from a mixed-layer mesoscale model); and duct heights derived from the satellite-IR duct technique.

1. ASSESSMENT AND PREDICTION

Exploiting electromagnetic propagation effects on Naval systems requires the ability to assess, as well as predict significant features of the tropospheric refractive environment over the sea, for regions of operational interest. In either case, once the valid present or future conditions are defined, a vertical profile of refractivity and specific evaporation duct parameters can be input to IREPS [1], along with specific system parameters, to compute such products as coverage displays, path-loss diagrams, propagation summaries, and airborne electronic warfare stationing aids.

For evaporation duct calculations, measurements of ship-board temperature, humidity, wind speed and corresponding sea water temperature are needed as prescribed by Paulus [2]. To determine the vertical profile of refractivity required in assessing the effects of elevated ducts, and surface ducts caused by elevated (or near surface) layers, measurements based on radiosonde, air-borne refractometer or dropsondes are customarily made. Of these, radiosondes launched by balloon from aboard ship, or nearby coastal/island stations, provide the bulk of the data available.

The principal assumptions involved in predicting significant refractive structure lie in the basic relationships between refractivity and atmospheric moisture, temperature and pressure. Since weather prediction models typically make forecasts for the latter three atmospheric parameters out to 3 to 4 days, it is assumed that refractivity itself should also be predictable.

The horizontal continuity of atmospheric layers and their relation to air masses is the logical basis of the assumption that refractive structure can be forecast. Indeed, Bean [3] was able a quarter of a century ago to map out various refractive parameters as a function of air mass conditions. By forecasting the positions of highs and lows depicted on routine weather maps, it follows that forecasts of derived fields, like refractivity, should be feasible. And it is, -- up to a point. We find, however, that in reality, the occurrence, height and intensity of ducts may at times significantly differ from forecasts made by the best forecasters.

What makes successful predictions of refractivity still so elusive are the following: (1) We are trying to forecast vertically small features by inferring, rather than measuring their existence from observation of large (synoptic) scale atmospheric conditions; (2) the occurrence, height and intensity of ducts are influenced not only by synoptic features, but also by terrain-induced mesoscale features, gravity waves and other perturbations; (3) the critical data (radiosondes, etc.) that are used to measure and validate the refractive conditions, as well as provide data for initialization of numerical prediction models are subject to some degree of uncertainty and require careful application; and (4) the prediction models used to forecast the future state of the atmosphere are themselves imperfect, as are all models.

One could of course, as a first guess, simply rely on climatological expectations of ducting conditions. Extensive efforts have been conducted by Ortenburger [4] to develop world-wide data bases which are included as library call-up options in IREPS for planning operations well in advance. But for tactical applications, actual forecasts remain a high-value item.

At the request of the Commander, Third Fleet, and in response to Pacific Missile Test Center (PMTC) concerns over the effects of refraction on missile test operations, an initial effort to develop a synoptic-refractive relationship for operational forecasting use was attempted by Rosenthal [5] resulting in the development of the Refractive Effects Guidebook (REG). The REG employed simple procedures for estimating present or future refractive conditions (and hence systems coverage) from information supplied by synoptic charts, transmitted Route Weather Forecast (WEAX) messages, or direct measurements. Relationships between refractivity profiles and surface weather patterns were deduced by a combination of subjective and objective techniques using a 1974 data base of radiosonde measurements and synoptic charts for mid-season months. Although the technique was found to be procedurally feasible and attractive by Fleet Units, objective evaluations of REG performance by Glevy and Logue [6] indicated significant discrepancies in verification, particularly for surface-based ducts.

Subsequent efforts by PMTC to develop acceptable operational guidance have concentrated on determining statistical relationships between synoptic weather variables and both duct occurrence and height. First, Rosenthal and Helvey [7] performed a comprehensive review of the literature to establish

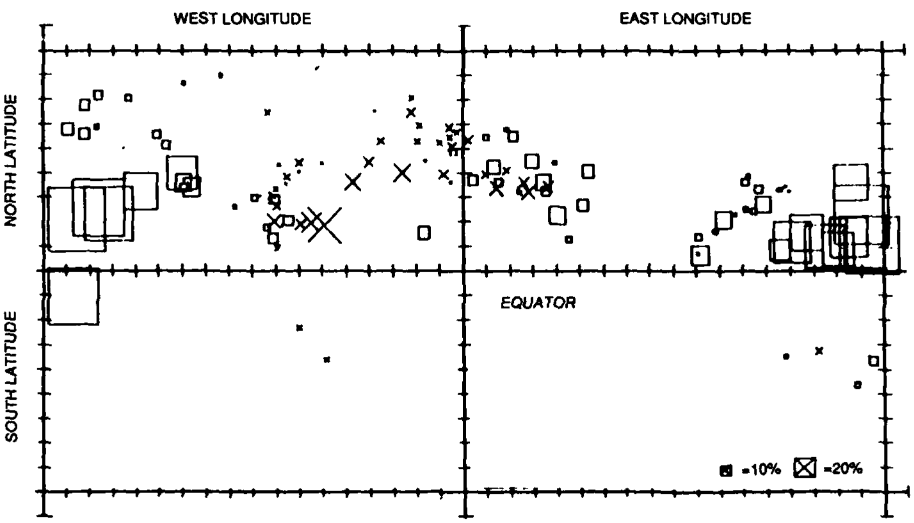
* (Now with Texas A&M University)

the conceptual basis of synoptic-refractive relationships and then developed statistical results using data from subtropical/temperate latitude radiosonde stations in the Pacific (Ship NAN and Midway Island) and in the Atlantic (Terceira, Ship H, and Bermuda). Additional studies were conducted by Helvey [8] and Helvey and Rosenthal [9] to examine frequencies of duct occurrences up to 10,000 feet for surface weather parameters using an expanded data base including the addition of Wake Island and over 3,000 soundings. These efforts included some case studies showing duct occurrence and height variations with synoptic weather conditions.

2. LIMITATIONS

Despite significant correlations found in these relationships in a general sense, an inherent remaining difficulty in these efforts is the necessity of inferring relatively small mesoscale phenomena such as refractive layers several hundred feet thick, from analyses of large scale air mass properties covering hundreds of miles defined by surface pressure patterns and analyzed air mass boundaries.

This task is made considerably more difficult by certain deficiencies in the available data base. In attempting to relate weather patterns to refractive structure over the ocean, one must use radiosonde data obtained from island and coastal locations. These locations are subject to localized wind, temperature and moisture variations induced by terrain which can add additional complicating structure. At the same time, the radiosonde measurements themselves are vulnerable to instrumental and procedural errors which indicate that much of the evidence for oceanic surface-based ducts (non-evaporative ducts) is erroneous as described by Helvey [10]. This discovery resulted from the observation that there were substantially more frequent occurrences of surface-based super-refractive and ducting layers noted in daytime than in night-time refractive climatological data derived from standard United States radiosondes, as shown in figure (1).



SYMBOLS ARE CENTERED ON RADIOSONDE SITES AND SIZES ARE PROPORTIONAL TO PERCENTAGE DIFFERENCE IN LAYER OCCURRENCE (SEE INSET)
□ MORE OCCURRENCES AT 0000 GMT x MORE OCCURRENCES AT 1200 GMT
DATA SOURCE IS 'COASTAL' STATIONS FROM 1972 GTE SYLVANIA REPORT, MAY 1966-APRIL 1969

FIGURE 1. Global distribution in relative frequency of occurrence of surface-based super-refractive layers. (from reference 10)

Inaccuracies in radiosonde data arise from many factors, including sensor characteristics, sonde design and materials, handling prior to release, and reduction procedures. Pressure, temperature and humidity from the sonde are required for computation of microwave refractive index, with humidity generally the most critical parameter. Humidities were found to be too dry in daytime, in U.S. radiosonde data, as a result of solar-radiation effects on the humidity sensor(hygristor), even after the introduction of improved sondes. The error arises because the heated hygristor warms and thereby depresses the relative humidity of the air being measured. For measurements within the first few hundred feet of ascent, procedures employed during pre-release preparation of the sondes can contribute to this dry bias. Even when reasonable care is exercised to avoid prolonged exposure to direct sunlight, appreciable warming of the instrument case and hygristor is probable, because of poor ventilation while the sonde is held stationary during the period just prior to release. Because of hygristor thermal lag, the effects of any initial temperature excess will persist for some distance above the surface, prolonging and increasing the magnitude of the refractive deficit aloft generated by ascent through the daytime surface super-adiabatic layer.

It is important to recognize and compensate for these effects if possible (as suggested by Helvey [10], [11] to avoid concealing or distorting the desired refractive observational/predictive relationships. Of course, other types of radiosondes will not behave in an identical manner, but depending on sensor and data sampling characteristics, may likewise impress their "personalities" on the derived refractive data. Radiosonde measurements are, nevertheless, the best and only source of data available to establish refractive climatologies and weather relationships on a world-wide bas.s.

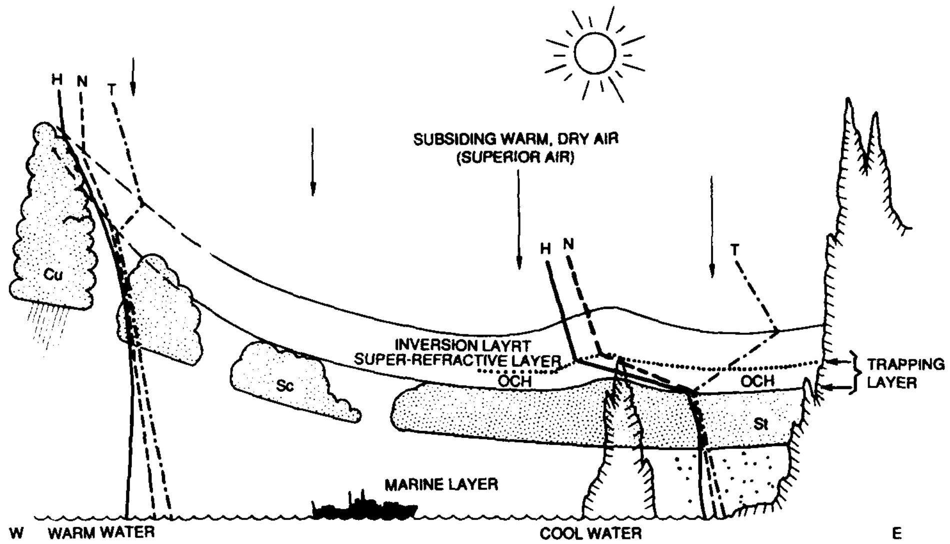


FIGURE 2. Schematic view of typical inversion conditions over subtropical oceans. (from reference 12) Inversion/super-refractive layer is lower and stronger over eastern side off ocean (with stratus (st) below) and is higher and weaker (with penetrating stratocumulus (sc) and cumulus (cu) in western side of ocean. A typical temperature (T), humidity (H), and refractivity (N) profile is shown at each side. Optimum coupling height (OCH) is near the cloud top.

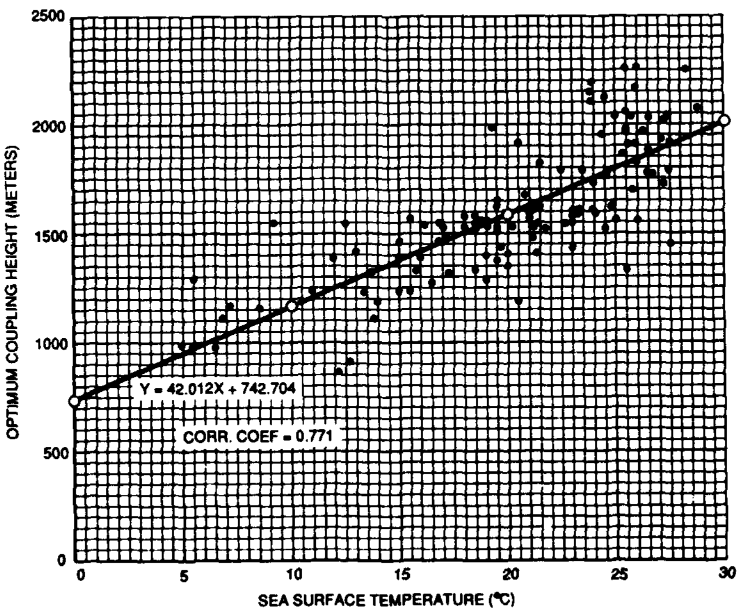


FIGURE 3. Median optimum coupling height (from five-year Sylvania study) versus mean sea surface temperature (best fit linear regression line shown by solid line). (from reference 12).

3. REFRACTIVE INFERENCES FROM SATELLITE CLOUD APPEARANCE

Meteorological satellite data can provide another essential input to the refractive assessment and forecast process, by virtue of its ability to simultaneously and repetitively observe conditions over vast regions of the open sea where surface observations and radiosonde data are largely nonexistent.

A satellite inference approach using cloud-pattern recognition was outlined by Rosenthal and Helvey [12] in a summary of observed synoptic-satellite-refractive relationships. Cloud types are known to be associated with various weather and stability conditions. For example, over the eastern portions of the sub-tropical Pacific and Atlantic Oceans and some areas of the Indian Ocean/Arabian Sea, extensive low stratus and stratocumulus clouds prevail in regions dominated by subsidence and resulting temperature inversions. Where waters are coolest (from wind-induced upwelling) and subsidence strongest, the clouds are lowest and more stratiform, and the subsidence inversions and corresponding refractive layers are also lowest and strongest. Where the waters are warmer, mixing increases the height and decreases the intensity of the inversion, and low clouds become deeper, higher, and more convective in a manner shown schematically in figure 2. The relationship between the sea/surface temperature and the Optimum Coupling Height (OCH), -- where ducting is most effective (near the inversion base and cloud-top), is also shown statistically in figure 3. Satellite data verifies these statistical relationships by continuously revealing low clouds that become progressively more lumpy, with cell size increasing as the water gets warmer, or as the atmosphere becomes more unstable (figure 4).

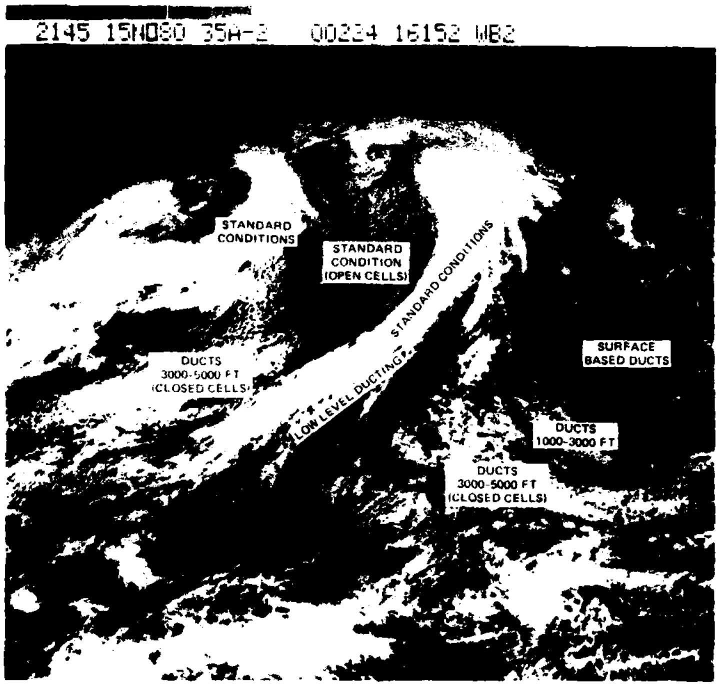


FIGURE 4. GOES imagery for 2145 GMT, 15 Nov. 1980, showing typical wintertime synoptic features over North Pacific superimposed by indications of probable refractive structure. (from reference 12)

At higher latitudes, or when cooler air masses invade these sub-tropical waters, frontal bands characterized by precipitation and strong winds are revealed on satellite imagery by long bands of thick layered and convective clouds. Within these regions, mixing is strong, ducting is absent and refractive conditions are near-standard. Behind these fronts, continued instability and standard propagation conditions are reflected in open-celled convective clouds. As stability returns, these cells become closed and once again signal the occurrence of super-refractive or ducting conditions. By recognizing the changes in appearance of satellite imagery cloud patterns, important information on duct occurrence and intensity can be inferred.

Even though qualitative analysis of satellite imagery cannot be expected to provide a precise fix on the altitude of ducts, it does have the potential for indicating duct occurrence, probable height categories, and trends of duct strength.

In an attempt to quantify and verify these observed relationships, a study was made by Rosenthal, Westerman and Helvey [13] of satellite-observed cloud appearance in the vicinity of San Nicolas Island (SNI), a Navy-owned island approximately 60 nautical miles off the Southern California coast at 33deg 15min N, 119deg 28min W and is the focal point for Navy missile testing at sea. The island is a regular site for rawinsonde measurements, and is therefore an excellent source of ground-truth for determining refractivity conditions. It is typically immersed in the northwesterly flow characteristic of this part of the Eastern North Pacific Ocean, making it relatively remote from major mainland modification of the marine layer except during periods of offshore flow. Some modification of marine air

flow conditions also occurs due to orographic influence by the island itself which has a maximum elevation of 907 feet. Nevertheless, since the elimination of Ocean Station Vessels NAN and PAPA, except for Hawaii, SNI is the most oceanward permanent site available off the West Coast from which adequate ground truth data are available.

Variations in stratus appearance were studied for cases of shallow, moderate, and deep marine layer conditions. From this came the observation that very smooth, structureless stratus (or clear areas adjacent to smooth stratus) were associated with very shallow marine layers and low, strong inversions; granular-looking stratus was associated with shallow to moderate marine layer depths and moderate to strong inversions; while progressively larger closed-cell stratus or stratocumulus was associated with deeper and deeper marine layers and higher and weaker inversions.

Satellite imagery used in this study consisted of GOES-West visual (0.55-0.75um) and infrared (10.5-12.5um) data received at up to 30-minute intervals at Point Mugu via a GOES-tap phone line hookup to the National Environmental Satellite Data Information Service (NESDIS) Satellite Service Field Station at Redwood City, California. While the satellite was positioned about 22,000 miles (36,000 km) above the equator at 135W longitude for the period of record, the sectorizing of the full-disc image allowed resolution of up to 0.5 mile at the image center for visual data and 5 miles for the infrared.

Satellite imagery was selected generally with respect to its proximity to the time of the closest sounding, however, when such imagery lacked the required resolution (such as large scale SB or UC infrared images), the nearest visual SA1 was usually selected. For the 1981 period of record, of the 326 images and sounding "pairs" used, 322 or 99% were within 2 hours of each other, with several within 2 minutes. On a few days, available imagery was either absent or inadequate to determine low cloud conditions and no correction was attempted. In general, only visual imagery provided the necessary resolution when using imagery received via GOES-tap phone hookup.

The appearance of cloud conditions close to SNI was observed for each image and the occurrence of stratus or stratocumulus, cell type and size (in tenths of a degree of latitude), proximity to fronts and other factors were tabulated. Actual duct occurrence, height, and intensity were independently determined from the rawinsonde nearest in time, and compared with the satellite cloud features.

Relationships between the (non-evaporation) duct height determined from the radiosonde data and cloud appearance were then formulated. General trends, consistent with prior expectations were observed with however, a significant amount of scatter, particularly for the higher ducts. Figure 5 shows typical examples of satellite cloud features, and the corresponding SNI sounding derived duct heights and intensity for the categories, "clear, near smooth stratus," "granular stratus," "small closed cells or dense stratus," "moderate closed cells or dense stratus," "large closed cells, stratocumulus," and "open cells, convective clouds." On each satellite image, the location of SNI is indicated by a small circle. Superimposed on each satellite image, according to the scale displayed to the left of the figure, is an indication of the mean and mean absolute deviations of duct base height for each of the major classification categories of cloud appearance.

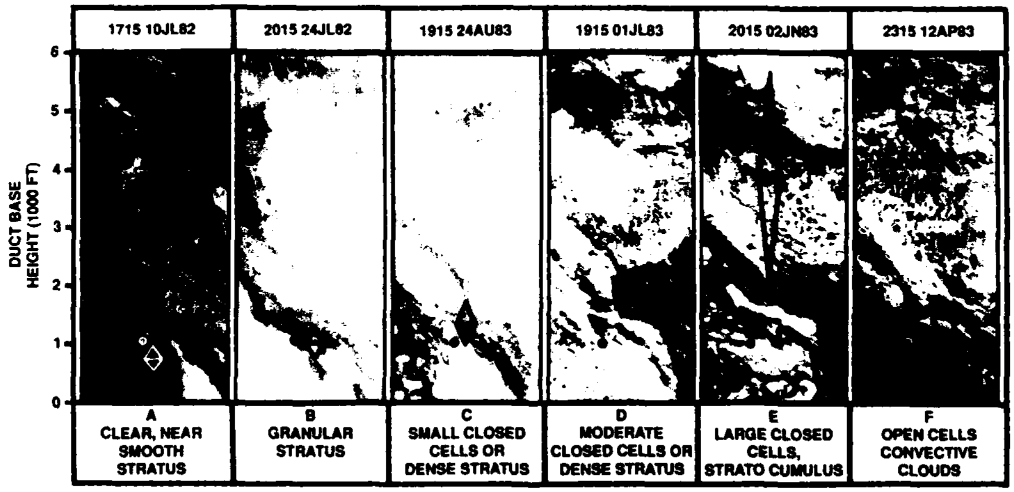


FIGURE 5. Satellite imagery patterns associated with refractive structure atSan Nicolas Island (SNI). SNI is located by circle. Duct base height-derived from statistical study is shown for each cloud category by diamond, showing mean and absolute deviations for height scale displayed on left.

As the appearance of stratus progresses from smooth, to granular, to larger and larger closed cellular (stratocumulus) clouds, the mean duct height (and its scatter) increases until ducts finally disappear with open convective cells.

In an attempt to evaluate the potential predictive value of using the results, various statistical verification scores were applied to validate the satellite-refractive relationships, when used to "predict" duct occurrence. These included percent correct of YES-NO forecasts (81.6%), power of detection (.94) and threat scores (.82).

Duct intensity was also a factor in the statistical findings. When ducts were forecast and did occur, intensities were fairly strong with about 41.2% of ducts being 30 or more M units in strength and 77.2% of ducts being 10 M units or more in strength. On the other hand, when ducts were forecast not to occur but nevertheless appeared on SNI rawinsondes, most of these were observed to be very weak and operationally insignificant. Only about one third were stronger than 10 M units, and only 5% were stronger than 30 M units.

This combined subjective/objective technique offers promise as a simple operational technique that Navy shipboard personnel can employ using passively received satellite imagery and a relatively minimal amount of training. While the technique is presently limited to mid-latitude or sub-tropical ocean areas, and precise duct heights cannot be determined, the technique can be useful in forecasting trends in duct height as the ship's course is compared with prevailing satellite imagery patterns. Moreover, the appearance of smooth, grainy or small-celled stratus can be used as a clue that low strong ducts are likely and that radiosonde measurements are needed for input into IREPS to more precisely locate duct altitudes and anticipated effects.

The information and thumb-rules derived from this study are being presently incorporated into an expert system so that the inherent information can be used interactively in a more objective and automated manner.

4. SATELLITE-DERIVED AUTOMATED DUCT HEIGHT ESTIMATES

A much more objective and quantitative approach to duct height assessment has been developed at PMTC by Lyons [14], which has great potential for application over open ocean areas, where radiosonde or other direct measurements aloft are lacking. As shown in figure 2, the tops of marine atmospheric boundary layer clouds which are present over much of the earth's ocean surface (such as the subtropical Northeast Pacific) are generally coincident with the base of a widespread subsidence inversion, and also the Optimum Coupling Height (OCH) of an associated elevated duct. Due to turbulent mixing, temperatures within and at the top of this layer are strongly related to temperature of, and altitude above the underlying sea surface, particularly in the presence of marine stratus/stratocumulus clouds. If sea surface temperature and cloud top temperature are known at some location, then cloud top/OCH altitude can be determined if the appropriate temperature-altitude relationship is also known.

The PMTC technique obtains cloud-top temperatures from infrared (10.5u-12.5u) geostationary satellite data, using an image display and processing system (SPADS) originally developed by Nagle, and described by Schramm, et al [15], at the Naval Environmental Prediction Research Facility, (NEPRF). Conversion from temperature to altitude (OCH) is accomplished through seasonal relationships established from statistical analysis of radiosonde profiles at San Nicolas Island, 60 miles offshore from Point Mugu, and adjusted according to the seasonal sea surface temperature difference between the point of interest (offshore Southern California) and San Nicolas Island. Because of meteorological/physical contraints, duct base, OCH and top are necessarily closely linked, with relatively small variations in vertical separation compared to the vertical domain of interest. Thus a constant (average) displacement can be used to determine approximate altitudes for duct base and top, based on the current OCH.

Data for Point Mugu, and Barking Sands (Kauai, Hawaii), have also been explored to establish seasonal temperature/height relationships. The former was found to behave similarly to San Nicolas Island, but with an offset comparable to differences in sea surface temperature between the two locations. Barking Sands is representative of the mid-Pacific region, and additional data from other sites and oceans is being analyzed to permit a generalization of the technique to any region of interest, where appropriate.

The accuracy of this technique depends on validity of the several relationships presumed or determined as outlined above. At present, cloud top temperature is the only measured, as opposed to statistically established parameter. Upper level moisture can be expected to cause cold biases (clouds appear too high) when present in sufficient amount, but attempts to correct for this effect by comparing satellite-observed sea surface temperatures against values derived from ground truth data have so far been unsatisfactory, due to occurrences of over-correction and problems in determining suitable reference locations. In areas with partial cloud coverage warm biases (clouds appear too low) will occur because the field of view for a given pixel may be contaminated by radiation from the sea surface; a correction procedure is being developed based on concurrent data at visual wavelengths, along the lines described by Chou et al [16].

Figure 6 shows the final product when displayed on the SPADS system. At each cloud-top point defined through use of the cursor, cloud-top temperature is displayed on the right and a set of three heights (in tens of feet) is displayed on the left: duct top, OCH, and duct base. The computed heights can be stored as an overlay and displayed atop either the source (IR) image or a near concurrent visual (0.55-0.75um) image.

5. VALIDATION AND MODIFICATION OF THE SATELLITE-IR DUCT TECHNIQUE

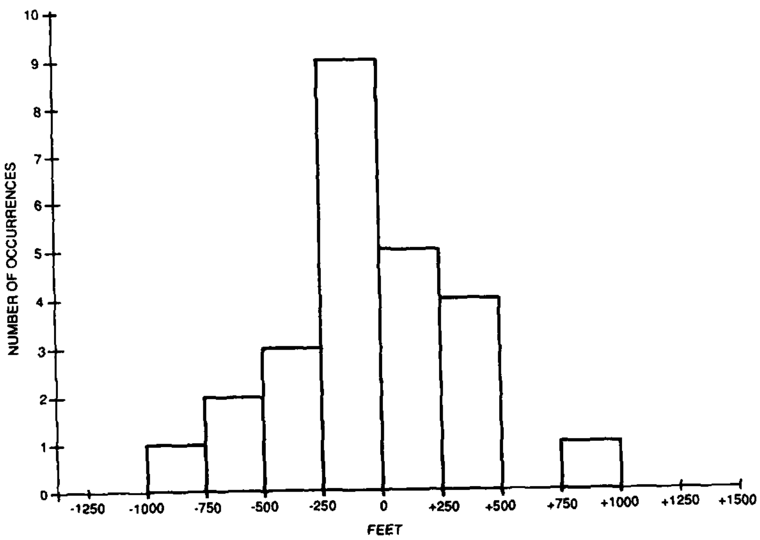
Since duct height estimates using this method would obviously be expected to be best closest to the Point Mugu/SNI locations, where data consistency was greatest, validation of the algorithms have so far been based principally on SNI sounding data. Initial evaluations of the technique for this area have been very promising. When comparing rawinsonde-derived OCHs with those generated by the Satellite-IR duct technique, Lyons [17] found correlation coefficients to be greater than .85 with about 80% of data points agreeing within 400 feet.

More recent evaluations by Szymber and Fox [18] have attempted to examine duct height consistency from the standpoint of reproducibility, as well as operator-to-operator differences. To address the former, triplets of nearby points (30-60 nmi maximum separation) were selected and compared to provide an indication of local variations due to meteorological differences between cloud elements, and noise inherent in the analog GOES-tap telephone signal input. Duct height estimates based on averages of closely spaced points has proven useful in reducing non-representative variability due to these small-scale factors.

To address the problem of operator differences, Szymber and Fox each independently derived OCH values from the same data set. Figures 7 shows a histogram of the distribution of RAOB-SPADS OCH height differences for one of these analysts. It should be noted that very high correlations (r=.947) occur when special care is exercised to select closely-spaced point sets for averaging, and avoiding points contaminated by sea surface or upper moisture effects.

In developing the technique, a sea surface temperature array (1-degree latitude-longitude spacing) covering the eastern and central Pacific was used to translate the seasonal profiles from SNI or Hawaii to regions remote from these locations. This SST array was inserted for the month of August, and while month to month variations is SST were expected and observed, it was assumed that consistency in the SST gradients would allow the August array to be satisfactorily used throughout the year. However, this was not validated initially.

FIGURE 6. PMTC satellite-IR duct height technique. Shown are 3 sets of values superimposed on IR image with duct top, OCH and duct base (in tens of feet) displayed to left, and cloud-top temperature to right of point.



ERROR INTERVAL IN SPADS ESTIMATED OCH VALUES (ERROR = SPADS OCH - RAOB OCH).

NUMBER OF CASES = 25
MEAN ABSOLUTE ERROR = 252 FEET
SPADS OCH WITHIN ± 300 FEET OF RAOB FOR 76% OF CASES
SPADS OCH WITHIN ± 400 FEET OF RAOB FOR 80% OF CASES

FIGURE 7. Histogram of error in satellite IR duct OCH values

In a recent effort, Szymber and Fox [19] examined the actual OCH height differences and sensitivities that would be observed by using the Satellite-IR duct technique with the August SST array for images obtained during other seasons. In a worst-case situation, satellite imagery from the month of February was used to highlight the different OCH values computed using a February SST array (inserted into the system) versus the original August array. Figure 8 shows this 13 February 1989 image with two sets of OCH values displayed for each point selected on the image.

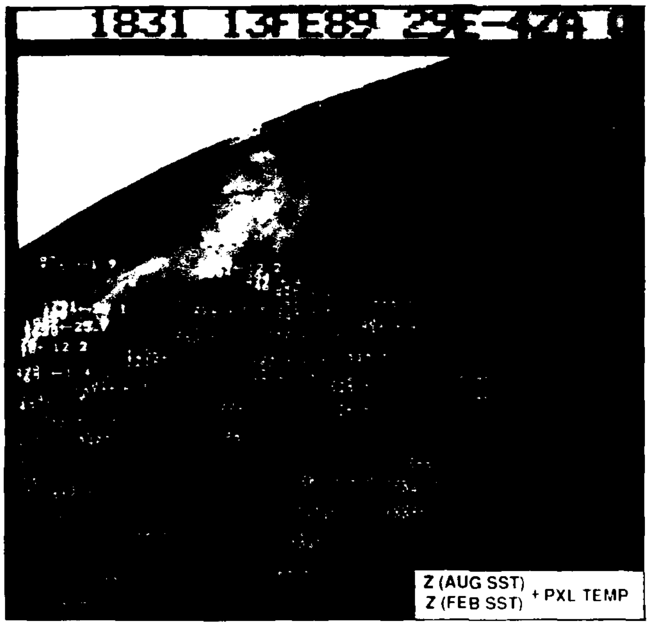


FIGURE 8. GOES IR image at 1831 GMT 13 February 1989 showing impact on OCH calculation using August (top) vs February (bottom) sea surface temperature array for each point selected. Number to right of each point is cloud-top temperature in degrees celsius.

Over the region between 40 to 50N and about 155W, OCH differences were as much as 3,000 feet. However, duct occurrence is climatologically very low in this region where the influence of mid-latitude systems is routinely felt. Over much of the east and east central Pacific south of 40 N, agreements were quite good with differences generally in the hundreds of feet. This indicates that for the sub-tropical regions where duct occurrence is high, and for which this technique is primarily intended, the impact of using an SST array for a month different from that of the satellite image is not too great. Nevertheless, the technique has been modified to use one of four seasonal SST arrays instead of the initial one.

To further check the sensitivity of computed OCHs to the use of an SST array for the same season but one month removed, such intraseasonal OCH errors were examined. Figure 9 shows an analysis of this error field for the Pacific region of interest for July OCHs computed using August vs July SSTs. The field is very flat with only a few spots where the error exceeds 500 feet. One of these is in the same temperate area north of 40 N as before where duct occurrence is minimal anyway.

Other validation efforts are planned as the technique is further refined for eventual inclusion in the Tactical Environmental Support System (TESS).

6. MEASURING AND ASSESSING HORIZONTAL INHOMOGENEITIES IN REFRACTIVE CONDITIONS

It is generally assumed that major refractive layers responsible for ducting conditions are horizontally stratified over the ranges within which naval shipboard and airborne radars typically operate. Previous NOSC studies by Gossard [20] and Glevy [21] provided evidence based on measurements that support this contention. At the same time, it is evident that in coastal regions, horizontal inhomogeneities in the atmosphere do exist. How large, frequent and significant these variations are, and to what extent, they exist over the open sea has become an issue of concern within the Navy's environmental community. Furthermore, even if substantial horizontal inhomogeneities are found to occur more frequently than previously thought, the question remains whether such atmospheric variations are significant to radar propagation itself. All this has a bearing on when, where and what propagation models need to be used for making EM propagation assessments.

Because of these unanswered questions, under the Battlegroup Environmental Enhancement project managed by NOSC, an effort is underway to address this issue. One problem that immediately surfaces is the very limited amount of conventional data available with which to document horizontal variations in the atmosphere at low levels over the sea. Because the satellite-IR duct technique permits duct height to be easily and quickly determined over large, Battlegroup-size regions of ocean covered by stratified low clouds, this technique is ideally suited to provide insights and data relevent to this question.

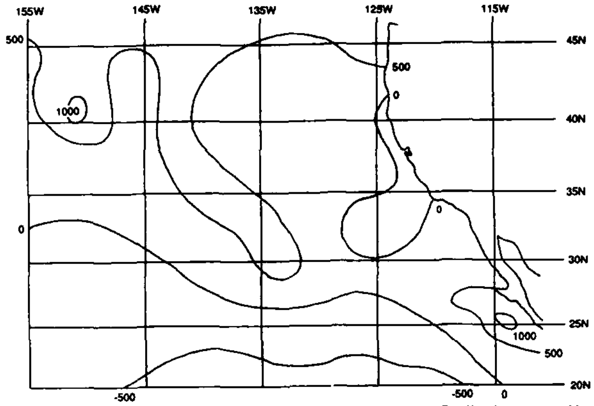


FIGURE 9. Intra-seasonal summertime errors in OCH calculation over Pacific Ocean resulting from applying the mean August SST field in July. July OCH error in feet.

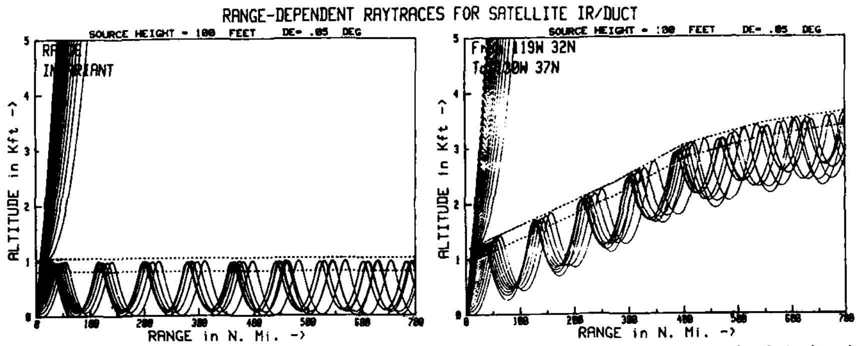


FIGURE 10. Range-dependent raytraces for range invariant condition (left) vs range variant condition (right) using duct values derived from satellite IR duct technique. Layer (level or sloping) is atmospheric trapping layer. Envelope of trapped rays also shown.

Figure 6 showed several duct heights determined from the satellite-IR technique along a line running from northwest to southeast across the Northeast Pacific Ocean. The estimated duct base height along this path varies from about 3300 feet to near 500 feet. If data were only available at one of those locations (e.g. point A) based on available radiosonde, dropsonde or climatological data, one would necessarily make the standard assumption that duct height was the same in all directions. However, the satellite technique permits the determination of additional duct heights in all directions from the chosen point as long as the requisite meteorological conditions (stratified low cloud field capped by subsidence inversion) and with accompanying elevated duct are met. In this way, any variability in duct height can be detected.

In order to qualitatively suggest how the satellite-determined variability in duct height in this example might affect EM propagation, a range-dependent raytrace developed by Helvey [22] for the HP-9020 desktop computer is used in figure 10 to show ray paths using the different satellite-determined duct heights as input.

If only the most southeastern set of data is used and no other information is available, horizontal uniformity is assumed and the ray trace in the left part of figure 10 indicates low near-surface based ducting would prevail with increasing range (to the northwest). But if all three data sets derived from the satellite technique are used, then the upward slope in duct height from about 32N 119W to 37N 130W and the corresponding raytrace is shown in the right part of figure 10. Even though raytracing techniques do not provide quantitative measures of propagation conditions, it is evident that inferred radar detection or communications conditions are significantly different when the atmospheric variability derived from the satellite IR-duct technique is included.

By repeatedly positioning the cursor on the satellite data processing system over a region of low cloud cover, approximate duct heights can be determined over a wide area. By analyzing the resulting field of data points (manually at present, but planned for automation later), the horizontal variability *of the atmosphere and duct topography can be mapped and examined statistically for geographical and other* relationships which can then be used for real time planning or the development of specialized climatologies.

For purposes of predicting refractive conditions over the sea, the same analyses of duct height can also be related to conventional synoptic weather map features through computer access to surface and upper air analyses from the Navy's Fleet Numerical Oceanography Center (FNOC) in Monterey, California. By overlaying the conventional weather fields directly on the satellite imagery (figures 11a/b), the variability in duct height can be related to proximity to high and low pressure systems, wind directions, temperatures aloft, fronts and other airmass properties. Since these fields (e.g. surface and 500 mb conditions) are standard products in numerical weather predictions out to 84 to 96 hours, documented relationships between satellite derived duct heights and synoptic features can be used to convert synoptic forecasts to predictions of duct height. Even for short term forecasting, ships in transit can potentially use either currently analyzed duct height conditions (now-casts) or 12 to 24-hour synoptic forecasts to infer duct heights at locations and times of subsequent arrival.

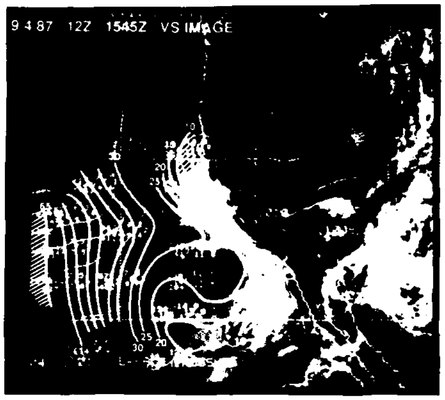


FIGURE 11A.

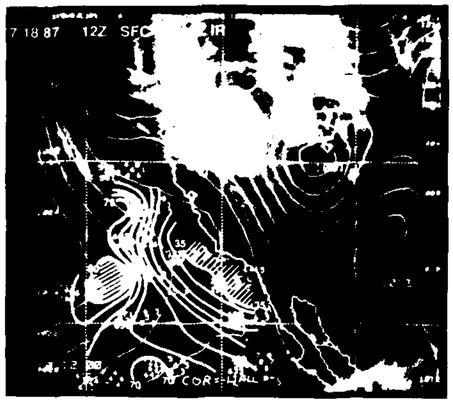


FIGURE 11B.

FIGURE 11. Array of OCH values displayed on GOES images. Superimposed are1200 GMT surface weather map patterns and manual analyses of duct height at intervals of 500 feet.
a. GOES visible image at 1545 GMT on 4 Sep 87
b. GOES IR image at 1315 GMT 18 Jul 87
Both show duct height increasing to southwest.

In Figures 11a and 11b, the duct topography was also hand-analyzed for two different summertime synoptic situations, the season of greatest duct occurrence in the Northeast Pacific. In Figure 11a, strong surface high pressure prevails off the U.S. West Coast with high pressure also extending in to the U.S. Pacific Northwest area. Along the coast from Northern California south to the desert regions surrounding the Sea of Cortez (Gulf of California), a thermally-induced trough of low pressure exists. These are typical summertime features in which subsidence and resulting low, strong inversions dominate the West Coast. Analyzed satellite-derived duct heights for this region show bases less than 1,000 feet over the entire coastal strip, but rising steadily to the west and southwest where duct bases near 35N 138W are at approximately 6,500 feet.

Figure 11b shows an example of a cool, unstable situation in which the thermal trough is absent and quite low pressure (< 1007 mb) covers the Pacific Northwest and western mountain states. These conditions resulted from a cold trough of low pressure at 500 mb with much deeper mixing, increased instability and reduced subsidence. Accordingly, the analyst of satellite-derived duct heights along the central and southern California coast revealed heights of around 3,500 feet, increasing west and southwestward to over 8,000 feet near 32N 129W.

It is interesting to note that while the synoptic situations were very different, and the duct heights correspondingly higher in the unstable situation, the total amount of NE to SW change in duct height was very similar. Even in other seasons, while marked changes in duct height distributions occur with varying synoptic regimes, the basic upward trend in duct height seaward from the coastline towards the west and southwest stands out as a prominent statistical feature. This persistent feature over the Eastern North Pacific agrees with measurements and analyses of the height of the sub-tropical inversion made by Neiburger [23] and Edinger [24]. It also follows the previously-noted trend for cloud cells to become bigger and higher in conjunction with an increase in sea surface temperature as exhibited in Figure 3. It seems evident that sea surface temperature, with warmer waters causing deeper mixing, and higher and weaker ducts, in both stable and unstable atmospheric conditions, is a major forcing function affecting the distribution of duct heights over the ocean. This observation should be useful for forecasting duct conditions because of the relative persistence of sea surface temperature conditions over the open sea.

It is also important to point out that some of the observed increase in duct height over the Eastern Pacific occurs in narrow zones, with significant implications for radio propagation conditions. The satellite IR-duct technique can be very useful for pointing out where these front-like features occur.

However, most of the observed increase in duct height over the Eastern and East-Central Pacific Ocean occurs more gradually, with height changes of only a few hundred feet per 100 miles. With such slow changes in duct height, it appears that the assumption of horizontal homogeneity over the open sea for typical radar ranges may be a good one.

Approximately 55 individual cases with satellite-derived duct analyses and overlayed synoptic fields have been compiled. These are being used to develop characteristic cloud/duct height patterns for various synoptic events, and will be converted into statistical summaries for use in expert predictive systems.

It is clear that the satellite-IR duct technique is a promising tool for diagnosing and predicting duct height and its horizontal variability. However, the significance and implications of such variability to actual EM system performance is being addressed quantitatively by Hitney at Naval Ocean Systems Center. It may well be that over the open sea, atmospheric variations are not strong enough to warrant discarding the assumption of horizontal homogeneity.

7. MODEL ASSESSMENTS OF MARINE LAYER DEPTH

Another approach by Eddington [25] being used at PMTC to measure atmospheric variability involves the use of a mixed-layer model, implemented on a Hewlett-Packard HP-9020 desktop computer to show the response of marine layer winds and thickness to coastal terrain. The model is a one-layer, two-dimensional, grid-point model. It assumes that potential temperature and wind are constant with height in the layer and that the layer is capped by an inversion. Effects of diabatic heating, water vapor, entrainment, and spatial variations of potential temperature are neglected in order to focus on topographic effects. Assuming an inversion-dominated regime typical of the Eastern North Pacific Ocean/U.S. West Coast region, terrain features are allowed to alter an otherwise horizontally uniform wind field. The model provides results showing variations in marine layer topography, wind direction and speed typical of features observed in sub-tropical coastal environments.

FIGURE 12. 3-D perspective display of marine layer height (meters) using mixed-layer model (reference 25).

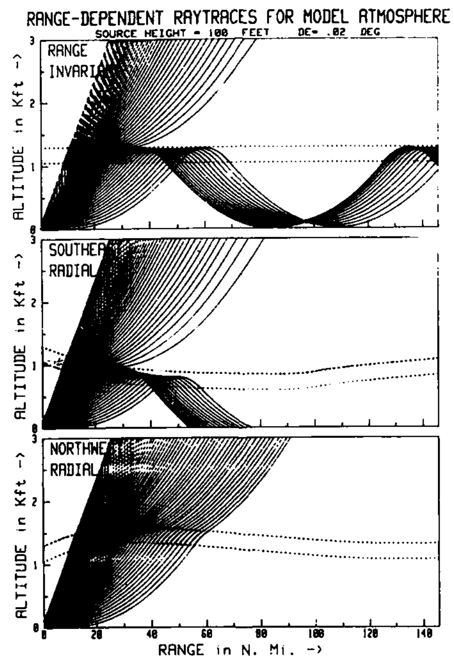


FIGURE 13. Range-dependent raytraces for range invariant condition (top), southeast radial (middle) and northwest radial (bottom) using data derived from mixed layer model solution.

Figure 12 shows a sample of a three-dimensional perspective diagram of marine layer height generated by the model and geographic terrain from an initially flat, (horizontally homogeneous) layer. Using typical wind directions and speed observed in the Southern California offshore area, the model 'predicts' height anomalies and differences of up to 2,300 feet between low points and high points. The derived structure seems realistic when compared to previous studies (e.g. Edinger [24] and recent day to day analyses. As was described for the satellite IR-duct technique, the model can also be useful in suggesting how EM propagation conditions may differ along various azimuths and ranges from some point source. Figure 13 shows range-dependent raytraces for 3 situations: the top one, assuming horizontally uniform conditions; the middle one showing how duct height (and suggested ray paths) change along the southeast radial; and the bottom showing changes along the northwest radial. While the quantitative impact on systems performance is not addressed, the model approach can be a valuable supplementary tool to the satellite-IR duct technique in diagnosing propagation conditions over marine environments.

8. FUTURE THRUSTS AND ONGOING WORK

In addition to the planned continued development, testing and refinement of the techniques described in this paper, several new areas are also being explored.

First of all, in order for the satellite-IR duct technique to be incorporated into TESS and be useful in the widest possible sense operationally, it must be usable in other ocean areas where similar conditions of inversion-capped stratus and stratocumulus occur. This includes regions off the west coast of South America, some areas of the Mediterranean and portions of the Atlantic off the coast of Africa, and the region of the Northern Arabian Sea just southeast of the Arabian Peninsula. Efforts are planned to use available radiosonde data to develop algorithms applicable to these regions.

Similarly, the mixed layer model has been used so far only for the terrain and geography of the region off Southern California. An attempt will be made to apply and test the model for other areas where the meteorological conditions are similar.

Since range-dependent raytraces are a convenient way to surmise some important effects on EM propagation from a horizontal-varying atmosphere as determined by either satellite or model output, several improvements are underway. These include incorporation in the raytrace of terrain profiles for specified azimuths from the source. For examination of raytrace sensitivity to various factors, the thickness, strength, altitude, and wave length, phase and amplitude of the idealized inversion or trapping layer topography can be specified as in Figures 14a/b/c/d. Thus, features implied by the model, but especially, those determined from inspection and measurement of cloud cell size and orientation determined from satellite imagery can be incorporated to make the raytrace displays more realistic. These in turn should improve the representativeness of any inferences made on how atmospheric variability can impact EM propagation.

Another improvement underway, is the development of an interactive technique whereby horizontal variations in the mixed layer height along specific azimuths determined from model displays are used to construct raytraces along these same paths.

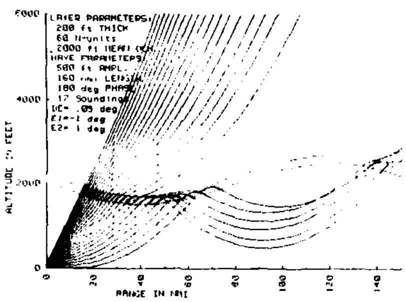


FIGURE 14A.

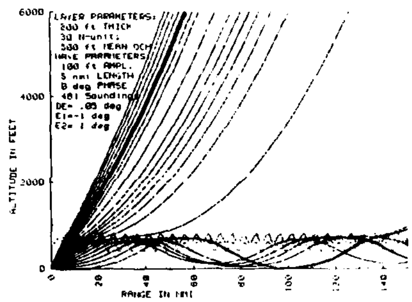


FIGURE 14C.

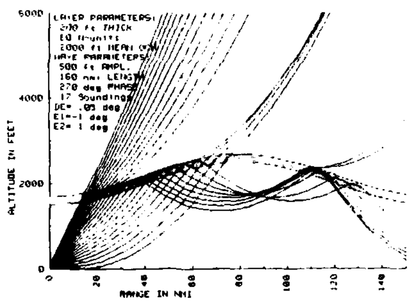


FIGURE 14B.

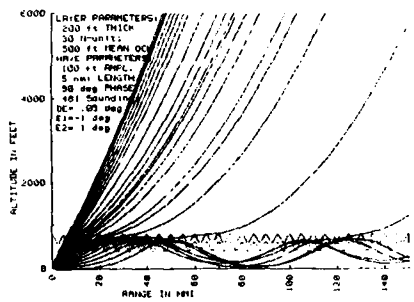


FIGURE 14D.

FIGURE 14A-D. Range-dependent raytraces showing varying wavelength and phase of wavy refractive layer. Input parameters shown. Wavelength used as input parameter could be derived from either *satellite-IR duct technique or mixed layer model.*

The satellite IR-duct technique, as indicated before, requires existence of low inversion-capped clouds for a duct height determination. There are, of course, many instances where strong subsidence intensifies and lowers ducts but also dissipates the stratus cloud layer. Some consideration has been given to the possibility of using water vapor imagery in the 6.7um band (Figure 15) to infer information about ducting conditions in such cloud-free areas.

A great deal of information can be surmised about atmospheric structure relevant to refractive conditions from the water vapor imagery (Rosenthal, et al [26]). When clear zones or holes on visual or standard IR imagery are present at the specific locations of interest, some of the early semi-empirical techniques based on inference of air mass regime from cloud patterns can be employed. Work is underway to convert these concepts and thumbrules into an expert system that can be used by personnel with minimal meteorological training. Such expert systems, in the form of operational decision aids, are also planned for implementation in TESS.

FIGURE 15. GOES water vapor image 2316 GMT 12 Jan 89. Dark regions are dry aloft and may infer regions of strong subsidence and ducting.

To address the needs of electro-optical and infrared systems, a low cloud/fog delineation technique also implemented on satellite data processing systems at PMTC will be used to highlight variability of cloud conditions at various altitudes that have a bearing on where energy might be totally absorbed by fog or haze partic s. As in the case of the satellite-IR duct technique, the low cloud delineation algorithms described by Fox [27] also rely on extraction of information from cloud top temperatures. An example is shown in Figure 16.

FIGURE 16. Low cloud delineation (reference 27) showing cloud tops derived from satellite image below 2000 feet, from 200 to 4400 feet, and between 4400 and 7000 feet.

Finally, work is continuing on the evaluation of radiosonde measurements since this data not only remains as the conventional source of information on refractive structure, but is also crucial to the evaluation of the other satellite and model techniques over the ocean. An "up-down" technique has been developed at PMTC, whereby soundings in both directions can be obtained from a single balloon launch to test the repeatability and spatial/temporal change of atmospheric structure. This technique has been successfully employed at several locations, including the Greenland/Barents Sea area in support of CEAREX-89. We plan to develop it as another viable tool to measure refractive structure and its variability, and validate inferences made from the satellite and model techniques.

REFERENCES

[1] W. L. Patterson, C. P. Hattan, H. V. Hitney, R. A. Paulus, K. D. Anderson, and G. E. Lindem, "IREPS 3.0 User's Manual," NOSC TD 1151, September 1987.

[2] R. A. Paulus, "Practical Applications of an Evaporation Duct Model," Radio Science, Vol 20, No. 4, pp. 887-896, Jul-Aug 1985.

[3] B. R. Bean, and E. J. Dutton, "Radio Meteorology," National Bureau of Standards Monograph 92, U.S. Government Printing Office, March 1966.

[4] L. N. Ortenburger, "Radiosonde Data Analysis (May 1966-April 1989)," Final Report. GTE, Sylvania, Electronic Systems Group, Western Division, Mt. View, California, 1973 (data base subsequently expanded).

[5] J. Rosenthal, "Refractive Effects Guidebook (REG)," Pacific Missile Test Center, Point Mugu, California, published by Naval Environmental Prediction Research Facility, Monterey, California, promulgated as COMTHIRDFLEET TACMEMO 280-1-76, 1976 (canceled in August 1978).

[6] LCDR D. F. Glevy, and AGC L. E. Logue, "An Evaluation of the Refractive Effects Guidebook (REG)," NELC TN 3268, 1976.

[7] J. Rosenthal, J. and R. A. Helvey, "Some Synoptic Considerations Relative to the Refractive Effects Guidebook (REG)," in Proceedings of Conference of Atmospheric Refractive Effects Assessment, NOSC Technical Document 260, 1979.

[8] R. A. Helvey, "Effect of Surface Parameters on Elevated Duct Occurrence and Altitude," Pacific Missile Test Center, Point Mugu, CA, Technical Publication TP-81-26, 1981.

[9] R. A. Helvey, and J. Rosenthal, "Prediction of Elevated Ducts Over North Pacific and North Atlantic Ocean Areas" -- Interim Report, 1981.

[10] R. A. Helvey, "Radiosonde errors and spurious surface-based ducts," IEE Proceedings, Vol- 130, Part F, No. 7, December 1983.

[11] R. A. Helvey, "Guidelines for Correction of Radiosonde-derived Refractive Profiles and Climatologies," Pacific Missile Test Center, Point Mugu, California, Technical Publication TP000002, 1982.

[12] R. A. Helvey, and J. Rosenthal, "Guide for Inferring Refractive Conditions From Synoptic Parameters," Pacific Missile Test Center, Point Mugu, California, Technical Publication TP-5, 1983.

[13] J. Rosenthal, S. Westerman and R. Helvey, "Inferring Refractivity Conditions from Satellite Imagery," Pacific Missile Test Center, Point Mugu, California, Geophysical Sciences Technical Note No. 96, 1985.

[14] S. W. Lyons, "Satellite-Derived Refractive Duct Height Estimates," Pacific Missile Test Center, Point Mugu, California, Geophysical Sciences Technical Note No. 98, 1985.

[15] W. Schramm, P. Zeleny, R. Nagle, and A. Weinstein, "The Navy SPADS, a Second Generation Environmental Display System," 9th Conference Weather Forecasting and Analysis, American Meteorological Society Preprint Volume, 1982.

[16] M. Chou, J. Childs and P. Dorian, "Cloud Cover Estimation Using Bi-spectral Satellite Measurements," J. Appl. Meteor., 25, 1280-1292, 1986.

[17] S. W. Lyons, "SPADS Automated Duct Height Statistics," Pacific Missile Test Center, Point Mugu, California, Geophysical Sciences Technical Note No. 100, 1985.

[18] R. Szymber, and A. Fox, "Verification of the PMTC Satellite-IR Duct Technique," Pacific Missile Test Center, Point Mugu, California, Geophysical Sciences Technical Note No. 146, 1989.

[19] R. J. Szymber and A. D. Fox, "Sensitivity of Satellite-Derived Refractive Duct Height Estimates to Sea Surface Temperature Variations," Geophysical Sciences Technical Note No. 147, Pacific Missile Test Center, Point Mugu, California, 1989.

[20] E. E. Gossard, "Refractive Index Variance and Its Height Distribution In Different Air Masses," Radio Science, Vol. 12, pp. 89-105, 1977.

[21] D. F. Glevy, "An Assessment of Radio Propagation Affected by Horizontal Changes In Refractivity," Naval Electronics Lab. Cen. Technical Note 3153, May 1976.

[22] R. A. Helvey, "A Range-Dependent Raytrace Capability," Pacific Missile Test Center, Point Mugu, California, Geophysical Sciences Technical Note No. 143, 1988.

[23] Morris Neiburger, David S. Johnson, and Chen Wu Chien, "Studies of the Structure of the Atmosphere Over the Eastern North Pacific Ocean In Summer, Part I, The Inversion Over the Eastern North Pacific Ocean," University of California Publications in Meteorology, Vol I, No. 1, 1961.

[24] J. G. Edinger, and M. G. Wurtele, "Marine Layer Over Sea Test Range," Pacific Missile Range, PMR-TP-71-2, 1971.

[25] L. W. Eddington, "Assessment of Coastal Marine Layer Anomalies Using a Mixed Layer Model," Proceedings of American Meteorological Society (AMS) Conference on Satellite and Coastal Zone Meteorology and Oceanography," Anaheim, CA, 31 Jan - 5 Feb 1988.

[26] J. Rosenthal and S. Lyons (In preparation), "Water Vapor Imagery - A Missing Link In Observing The Atmosphere"

[27] A. Fox, "Satellite Low Cloud Delineation Algorithm," Geophysical Sciences Tech Note No. 141, Pacific Missile Test Center, Point Mugu, California, 1988.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the considerable efforts of Mrs. Marilyn Galyon, Geophysics Division, and Mr. Tony Yost, Visual Communications Division, for preparation of the manuscript.

TROPO: A TECHNIQUE FOR PREDICTING PROPAGATION AND MODEM PERFORMANCE OF DIGITAL TROPOSCATTER SYSTEMS

by
Dr. Steen A. Parl
SIGNATRON, Inc.
110 Hartwell Avenue
Lexington, MA 02173-3198
USA

INTRODUCTION

In this paper we present the capabilities of a PC-based decision aid for digital troposcatter link engineering. The program includes propagation effects such as multipath and spatial diversity, as well as adaptive modem performance, transmitter emission control filtering, etc. The model uses the existing NBS method for predicting long term effects and for short-term effect below 1 GHz. Above 1 GHz it uses a turbulent volume scattering model. It uses numerical integration whenever possible to avoid coarse analytical approximations. A key function is its ability to evaluate advanced diversity receiver and modem techniques. A predecessor of this program has been used by the USAF to evaluate and engineer current and planned digital troposcatter links.

OBJECTIVE

The objective of this software program is to provide a tool for troposcatter link design, diversity design, and for analysis of existing links. The program is intended to be as precise as possible in the link and modem characterization as well as in the numerical algorithms used. It is felt that this will reduce a significant part of the uncertainty in link prediction. There is some evidence that the model provides reliable predictions with less margin for prediction errors. This margin is usually determined by the service probability, as defined in NBS Tech. Note 101.

Figure 1 shows a block diagram of a troposcatter communications system and also the functions modeled in the program.

BACKGROUND

TROPO/PC™ has grown out of the need for improvement over older models in several areas: better predictions than offered by NBS, incorporation of multipath effects and use of the turbulent volume scattering mode which has been the accepted theory for decades but was not used in the NBS method. That method depended on empirical measurements mostly below 1 GHz and did not separate different propagation modes (diffraction, random layer reflection, volume scatter). In a program for the US Army, SIGNATRON, Inc. developed a first such model [Monsen and Parl, 1980]. In a later effort for the Defense Communications Agency, the model was revised and expanded to include diffraction, long term effects, and modem prediction for the MD918/GRC and AN/TRC-170 modems [SIGNATRON, 1983]. This model has been widely distributed by the Defense Communications Engineering Center (DCEC) to the DoD community and it is currently being used by several US agencies. It is referred to as TROPO v.1.0. Another, even more precise model, was developed for RADC in 1984 [Parl and Malaga, 1984]. This model, however, only predicted short term effects and was intended only as a tool for theoretical evaluations to be used on a mainframe computer.

Over the years, several problems have been found with special uses of the TROPO v.1.0 model. It was also cumbersome to use since both link and modem prediction were integrated and a mainframe computer was required. It was decided to develop a new program that was more modular and avoided the numerical problems sometimes encountered in version 1.0. TROPO/PC™ is the result of that effort. It consists of the propagation program TPROP and the modem program TROMOD. While the tropscatter pathloss prediction is almost the same as in the previous model, the modem performance is implemented by a new program. It is capable of evaluating any Decision Feedback Equalizer Diversity (DFED) modem, including those currently available.

PROPAGATION MODEL

The propagation model consists of three key parts: short term effects (Rayleigh Fading), long-term effect (log-normal average received power), and the prediction uncertainty, a confidence level, measured by the service probability defined in NBS-TN-101. We first address the short term effects.

A large number of theories have been advanced to explain the observed behavior of microwave signals received far beyond the horizon. Both theory and practice started developing around 1950 when Booker and Gordon [1950] explained the signals by scattering from turbulence. A number of different turbulence theories were proposed [Megaw, 1950; Villars and Weisskopf, 1954]. The theory that has received the widest acceptance today is the turbulence theory of Obukhov [1941] and Kolmogorov [1943], based on research first reported in 1941. This theory predicts that the path loss depends on a wavelength as $\lambda^{5/3}$ (isotropic antennas) and on the scattering angle as

$\theta^{-11/3}$ (pencil beam antennas). The theory is based on single scattering from locally isotropic turbulence and has been validated experimentally.

Based on this theory, the received power is modelled by:

$$P_R = P_T G_T G_R \frac{\pi k^2}{2} \int_V d^3\underline{r} \frac{|g_T(\underline{r})|^2 \; |g_R(\underline{r})|^2}{R_t^2(\underline{r}) \; R_R^2(\underline{r})} \Phi_n(2k \sin\theta/2) \tag{1}$$

$$\triangleq \int_V d^3\underline{r} \; H(\underline{r})$$

where

V	is the scattering volume defined by the horizon and the antenna patterns.
P_R, P_T	are the received and transmitter power levels.
G_t, G_r	are the transmitter and receiver antenna gains.
g_T, g_R	are the normalized antenna patterns relative to boresight of the transmitter and receiver antennas in the direction of the point $\underline{r}$ of scattering volume.
R_T, R_R	are the distances from the point $\underline{r}$ to the transmitter and receiver antennas.
$k = 2\pi/\lambda = 2\pi f/c$	is the wavenumber.
θ	is the scattering angle, i.e., the angle between the lines from the transmitter and receiver terminals to the point $\underline{r}$.
$\Phi_n(\underline{k})$	is the wavenumber spectrum of the turbulence, or the three dimensional Fourier transform of the spatial correlation function of the refractive index fluctuations. It is assumed to be isotropic and only dependent on the magnitude, $\cong 0.033\, C_N^2\, \lvert\underline{k}\rvert^{-m}$. $m=11/3$ for turbulent scatter and $m=5$ for the empirical NBS model used below 1 GHz.
C_N^2	is the structure constant of the refractive index. The following model is used as default $C_N^2 = 8.14 \cdot 10^{-3}(m+1)/(m-3)/32 \cdot h^{-1/3} e^{-h/3200}$ and h is the height above ground in meters. This formula makes it possible to transition between the NBS model ($m=5$) and turbulent scatter ($m=11/3$).

Figure 2 shows the geometry of a tropo link, indicating some of the key parameters needed as input. The integration over the common volume is done so that the dependence of power vs. delay is determined. The result is the Power Impulse Response (PIR) function, $Q(\tau)$, which is key to evaluating the performance of a digital modem. The integration can also be used to predict the delay spread. Based on Taylor's frozen turbulence model, the moments of the Doppler spectrum are given by

$$\mu_h = \int_V d^3\underline{r} \; H(\underline{r})[(\underline{e}_T - \underline{e}_R) \cdot \underline{u}/\lambda]^n$$

where $H(\underline{r})$ is the total integrand in (1), $\underline{e}_T$ and $\underline{e}_R$ are direction vectors of the incident and scattered waves, and $\underline{u}$ is the wind velocity vector. However, due to the absence of reliable models of the wind, this has not been implemented in TROPO/PC™.

The integration in (1) is performed not only for each diversity, but also for pairs of diversities that may be correlated. The correlation coefficients are also determined as a function of delay. The result is a PIR matrix ($\underline{Q}(\tau)$) which is saved in a file for use by the modem program (TROMOD).

By performing a numerical integration over the common volume, the predicted path loss includes the so called "aperture-to-medium coupling loss." This loss is mostly a function of the model used and is often approximated quite coarsely. With TROPO/PC™ one need not be concerned about the exact value of this loss since it is implicit in the result. However, for reference to other technique, this loss is calculated by comparing the results of the integration with a theoretical model for wide beamwidths.

The $C_N{}^2$ profile used is representative of the worst month (winter afternoon, dry air). To determine the long term deviations from this "reference point" the NBS TN101 or MIL-HBK-417 models may be used. They model the climate variations and provide estimates of deviations from the median reference path loss and the standard deviations from the median. The key input parameters are the effective terminal heights above the terrain out to the horizon. An "effective distance" parameter is calculated and used in the NBS approach to determine the parameters over the long-term log-normal distribution. The NBS model, and its derivatives, are used because of a lack of better atmospheric characterization, and it is based on a large database. Unfortunately, these models do not distinguish between troposcatter and diffraction components, and a large part of the predicted variation is due to the influence of diffraction. This can lead to larger predicted variations for the troposcatter components than occur in real life and hence to pessimistic results since diffraction usually does not contribute to outages with modern digital modems.

Another problem is that long-term variations of delay spread can not be determined from the NBS data. While this is alleviated by the robustness of newer modems in the presence of multipath spread, it shows that a better model is needed. Clearly the long-term variations in RSL and delay spread are primarily due to variations in the mean refractive index gradient and the refractive index variance ($\sim C_N^2$). Gossard (1977) has provided a model of refractive index variance that could be used for long-term calculations. Atmospheric refractivity data (Bean et al., 1966) can also be used but these approaches have not been validated with real troposcatter data.

The third aspect to address is the prediction reliability. Rice et al., (1967) define a service probability for that purpose. They use empirical data to estimate an increased path loss variance to meet a specified service probability. We expect that the present model is more accurate since it avoids use of approximations, requires detailed link characterization, and is based on a newer scattering theory. The database is too small, however, to estimate the prediction reliability at this point. In the next section, we show comparison with measurements that may help illuminate this point.

Comparison with Propagation Measurements

In a recent effort by the USAF, measurements of angle diversity signal levels were obtained on an L-band link and an S-band link (Parl, 1988). Previous measurements are available at C-band (Sherwood and Suzemoto, 1976). We now compare predictions with results from these measurements.

Figure 3 shows the measured and predicted results for L-band. Data were taken only for the month of December, 1986. This month is representative of the worth month conditions. The 95% prediction is clearly very conservative. The 50% prediction matches the data well, except the predicted variance is larger than measured and the measured RSL falls below the predicted for two channels in the 1%-10% range. This is believed to be due to the small data base (one month) and partial equipment outages.

Figure 4 shows the S-band measurements and prediction. Again, the 50% prediction is quite accurate while the 95% prediction is pessimistic. At around the 1% point a transmitter outage causes the measured distribution to fall below the predicted. The variance of the log-normally distributed RSL is smaller than predicted also in this case.

Figure 5 shows the prediction for C-band on a short link with 8' antennas. The RSL measured by Sherwood and Suyemoto [1978] is 6-8 dB below the predicted. With 15' antennas (Figure 6) the measured data agree quite well with the predictions. The same holds for a longer link with 28' antennas (Figure 7). This confirms our belief that the main *uncertainty* in current predictions is the atmospheric modeling, both the variation of C_N^2 vs height and the standard deviation. In Figures 5-7 the measured standard deviation is larger than predicted.

Modem Performance

The modem performance in the model is implemented in a separate program TROMOD. This program takes two files as inputs. One is called PROFILE.DAT, specifying the diversity configuration assumed, the power impulse response, and the long term means and standard deviation of the carrier-to-noise ratio. The other file, TROMOD.INP specifies the modem and custom modem parameters, the diversity configuration desired, bandwidth, data rate, code rate, etc. PROFILE.DAT can be generated by the propagation program TPROP or created by an editor.

Some of the key issues in the modem evaluation are spectral efficiency, performance criteria, demodulation process, and the use of coding. We discuss these issues next.

Spectral efficiency is the ratio of the data rate to the allocated bandwidth. It depends on the pulse waveform and the bandwidth constraint. The program currently allows three waveforms: 1) square wave (used in MD918), 2) raised cosine pulse (used in S575), and 3) raised cosine spectrum (used in MD1208 specification). It accepts four bandwidth constraints: 1) no constraint, 2) 99% bandwidth specified, 3) FCC19311 spectral mask, and 4) Shape Technical Center recommended spectral mask. For example, the user can specify a rectangular pulse, 7 MHz bandwidth with the STC mask and 12 Mbit/s. The program determines that a three pole Butterworth filter with 5.6 MHz bandwidth is required, and that the transmitted waveform has a peak-to-average ratio of 2.9 dB. Figure 8 shows the mask and the calculated spectrum. The program reduces the transmitted power accordingly and will evaluate the performance, taking into account the Inter-Symbol Interference (ISI) introduced by the transmitter filter. A receive filter will also be taken into account. Such a filter is usually required to reduce adjacent channel interference and limit the noise bandwidth.

The program calculates two short term performance measures as a function of E_b/N_0. One is bit error rate averaged over Rayleigh fading. The other is the probability of fading below a level where the bit error rate on a non-fading channel exceeds a specified threshold. This second measure is called the fade outage probability. TROPO v.1.0 also evaluated fade outage per call-minute and 1000-bit block

error rate, in accordance with DCEC 12-76 (Kirk and Osterholz, 1976). These are not implemented in TROPO/PC", both in the interest of simplification and to avoid approximations. In TROPO v.1.0 these measures were implemented with highly simplified assumptions about fade rate. The draft MIL-STD-188-323 recommends error-free second probability for a 64 kbit/s channel as the quality measure. It may be implemented in the near future assuming slow fading on the channel. Coding gain is taken into account with the outage probability, but not for the average bit error rate calculation.

Based on the short term performance and the specified mean CNR and the standard deviation of the RSL, the performance over the long-term, log-normal distribution is evaluated by numerical integration. This results in a yearly average bit error rate and a yearly fade outage probability. For voice channels, outage probability is a better measure of quality and it also reflects the real performance improvement achieved by using coding. Outage probability is also better for most data channels requiring retransmission if errors are detected. Error free seconds performance is closely related to outage probability.

The demodulation process assumes a decision feedback equalizer consisting of a matched filter, a forward equalizer, a backward equalizer, a detector, and, optionally a decoder. Figure 9 shows the general structure. The forward and backward filters are transversal filters and the number of taps can be specified in the input file, as can the delay of each tap. Defaults for these parameters exist for each implemented modem. The performance evaluation assumes QPSK but uses the more conservative DPSK error probability. The difference between coherent QPSK and DPSK allows for an implementation margin which is usually adequate to characterize real modem performance degradation. Additional degradation can be specified at the input, but the default, modem dependent value is recommended. The performance accounts for the residual ISI that the equalizer cannot handle in the way described by Monsen [1973].

As an example of what the program can calculate, Figure 10 shows the outage probability with a 10^{-4} threshold for two modems, one with three taps on the forward equalizer (MD918/GRC), and one with six taps (S575). At large delay spreads, both modems degrade, but the modem with only three taps is clearly much more sensitive to delay spread. Figure 11 compares angle diversity with space-polarization diversity. At moderate delay spreads, angle diversity is superior, but it degrades more rapidly at large delay spreads. This is due to the larger multipath spreads in the upper angle diversity beam.

CONCLUSION

We have described a digital troposcatter performance model that is the latest in a series of models developed ove the last ten years. The new model is simplified in several respects but has a number of new features necessary to aid today's troposcatter link engineering. The model has been validated with a limited set of measurements.

REFERENCES

Bean, B.R., B.A. Cahoon, C.A. Samson, and G.D. Thayer. A World Atlas of Atmospheric Radio Refractivity, ESSA Monograph 1. Washington, DC. U.S. Government Printing Office, 1966.

Booker, H.G. and W.E. Gordon, "A Theory of Radio Scattering in the Troposphere", Proc. IRE, Vol. 38, 1950, pp. 401-402.

Draft MIL-STD-188-323, Defense Communications Engineering Center. "System Design and Engineering Standards for Long Haul Digital Transmission System Performance", November 1984.

Gossard, E.E., "Refractive Index Variance and Its Height Distribution in Different Air Masses", Radio Science, Vol. 12, No.1, 1977, pp 89-105.

Kirk, K.W. and J.L. Osterholz, Defense Communications Engineering Center. "DCS Digital Transmissin System Performance", 1976, DCEC TR-12-76.

Kolmogorov, A.N., "Dissipation of Energy in Locally Isotropic Turbulence", Compte Rendus (Doklady) l'Acadamie des Sciences de l'USSR, Vol. 32, No. 1, 1943, pp 16-18.

Megaw, E.C.S., "Scattering of Electromagnetic Waves by Atmospheric Turbulence", Nature, Vol. 166, 1950, pp 1100-1104.

Military Handbook, "Facility Design for Tropospheric Scatter (Transhorizon Microwave System Design)", November 1977, MIL-HDBK-417.

Monsen, P. "Theoretical and Presumed Performance of a DFE Modem on a Fading Multipath Channel", IEEE Trans. Comm., V-COM-25, No. 10, 1977, pp 1144-1152.

Monsen, P. and S. Parl, SIGNATRON, Inc. "Adaptive Antenna Control (AAC) Program" 1980, Final Report CSA-76-8085-6.

Monsen, P., S. Parl, and A. Malaga, SIGNATRON, Inc. "Digital Troposcatter Performance Model", 1983, Final Report, Users Manual, and Software Documentation.

Obukhov, A.M., "On the Distribution of Energy in the Spectrum of Turbulent Flow", Bull. Acad. Sci USSR Geo. and Geophys., Ser. 4-5, 1941, p 453.

Parl, S. and A. Malaga, SIGNATRON, Inc. "Theoretical Analysis of Microwave Propagation", 1984, Final Report RADC-TR-84-74.

Parl, S., SIGNATRON, Inc. "Angle Diversity for DCS Troposcatter Program", August 1988, Final Report for Scientific Systems, Inc. and Rome Air Development Center.

Rice, P.L., A.G. Longley, K.A. Norton, and A.P. Barsis, National Bureau of Standards. "Transmission Loss Predictions for Troposcpheric COmmunication Circuits", 1967, Revised NBS Technical Note 101.

Sherwood, A. and I. Suyemoto, MITRE Corp. "Multipath Measurements Over Troposcatter Paths", 1977, Technical Report ESC-TR-77-252.

Tatarskii, V.I., National Technical Information Service. "The Effects of the Turbulent Atmosphere on Wave Propagation", 1971.

Villars, F. and V.F. Weisskopf, "On the Scattering of Radio Waves by Turbulent Fluctuations of the Atmosphere", Proc. IRE, Vol. 43, No. 10, 1955, pp 1232-1239.

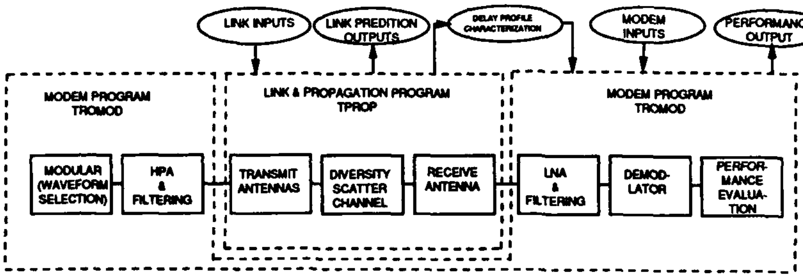


Figure 1 Functions Implemented by TROPO/PC™

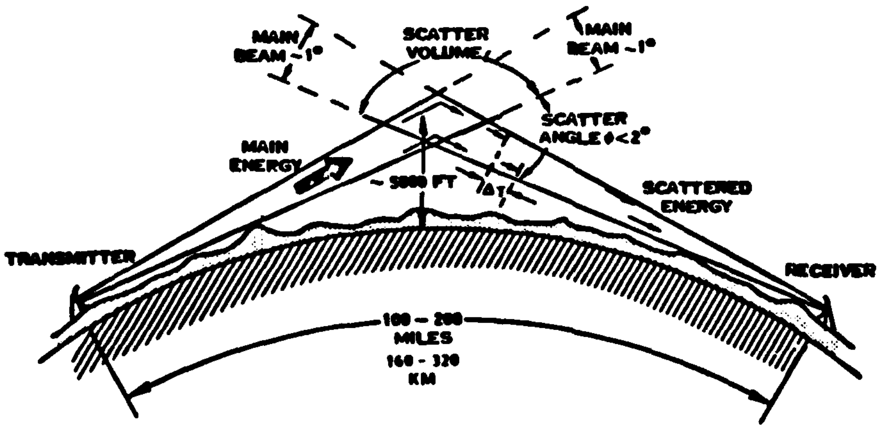


Figure 2 Path Geometry

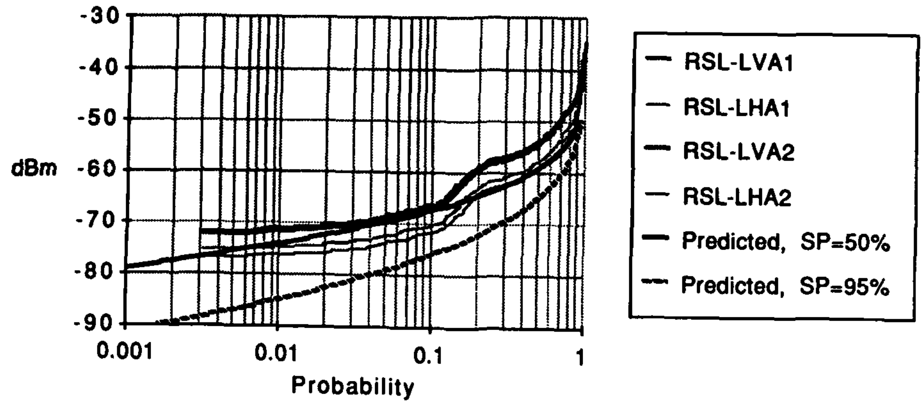


Figure 3 RSL Distribution of all data at Elmadag

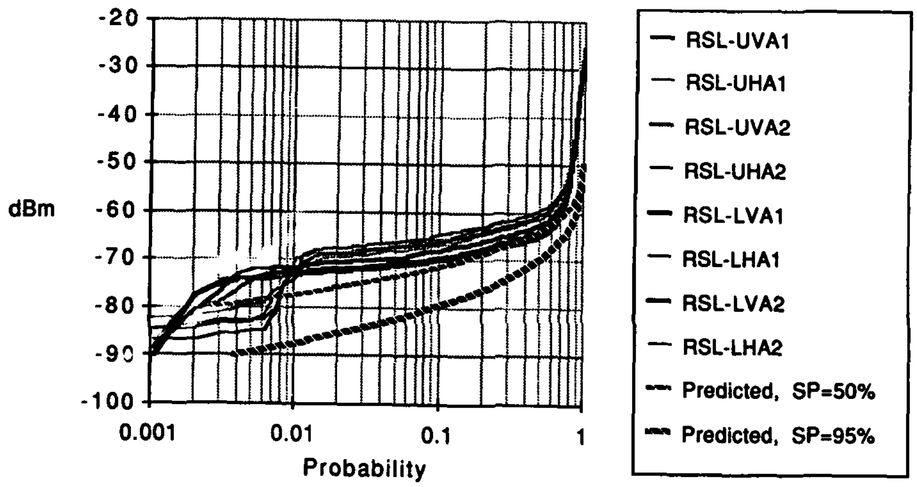


Figure 4 RSL Distribution of all data at Mt Limbara

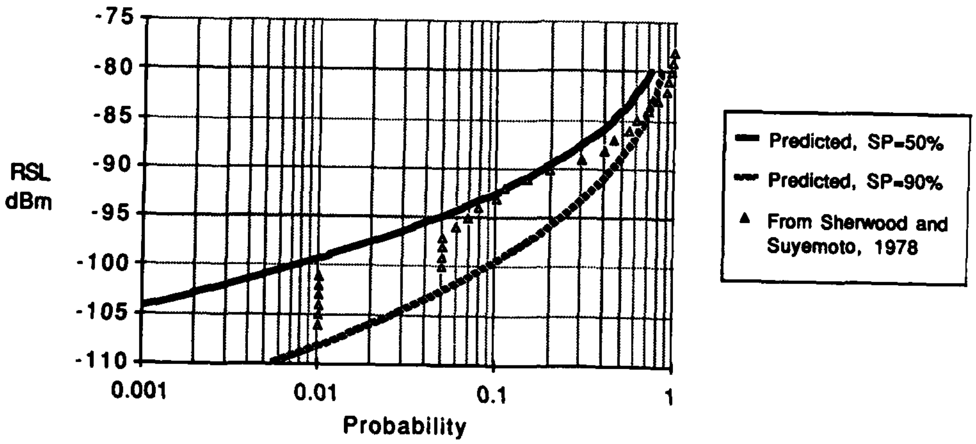


Figure 5 Ontario, NY-Verona, MY 5 GHz; 15' antennas

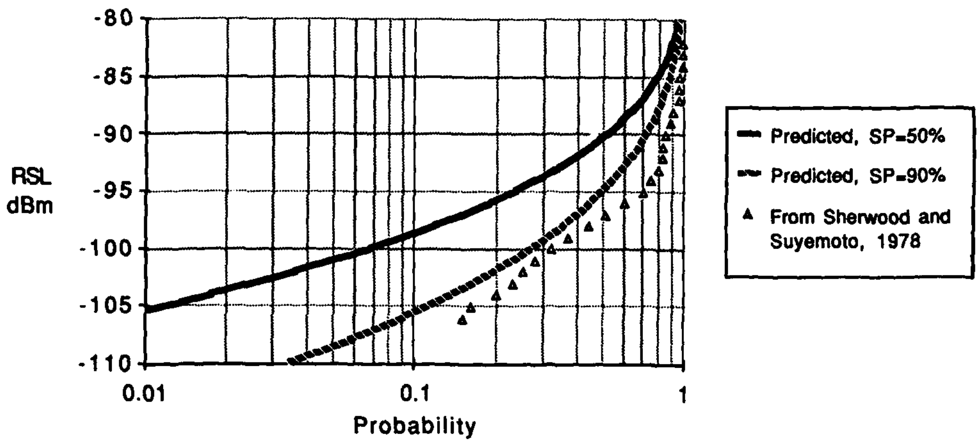


Figure 6 Ontario, NY-Verona, NY 5 GHz; 8' antennas

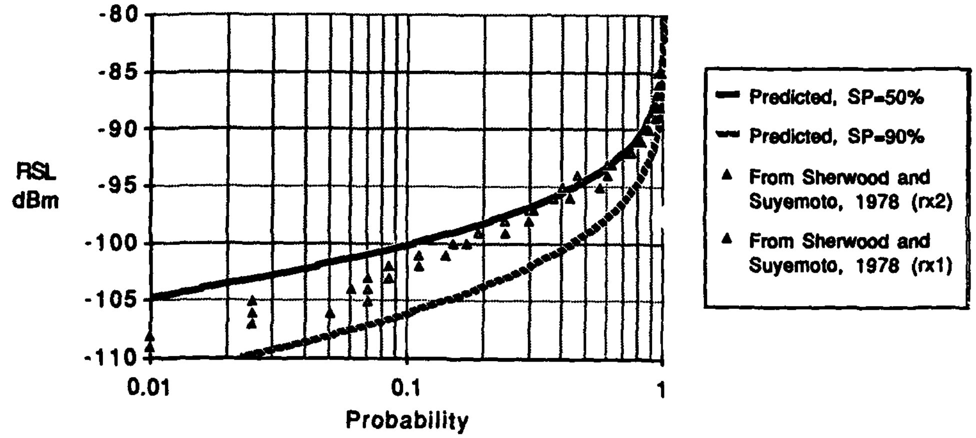


Figure 7 Youngstown, NY-Verona, NY 5 GHz; 28' antennas

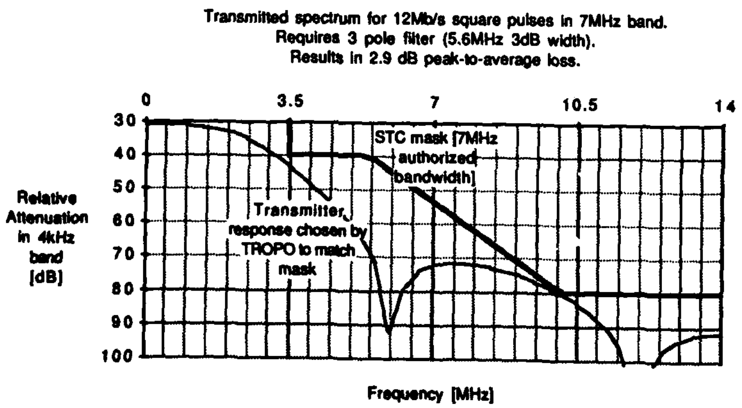


Figure 8 Transmitted Spectrum and Mask

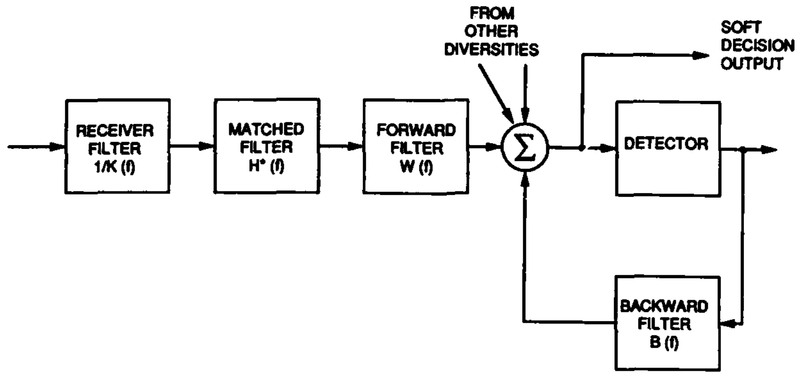


Figure 9 Decision-feedback Equalizer Receiver

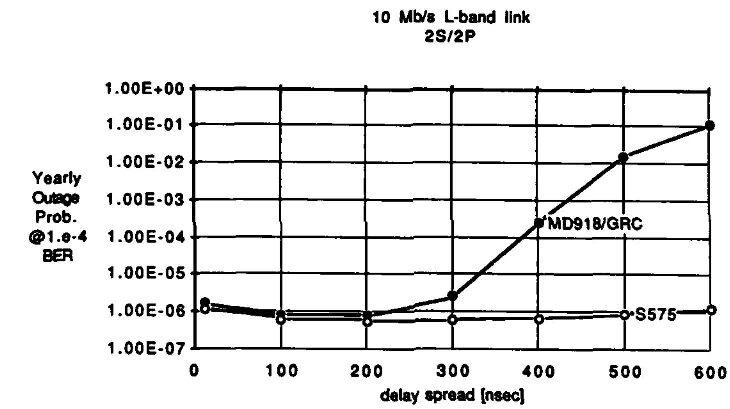


Figure 10 Lband 2S/2P Outage

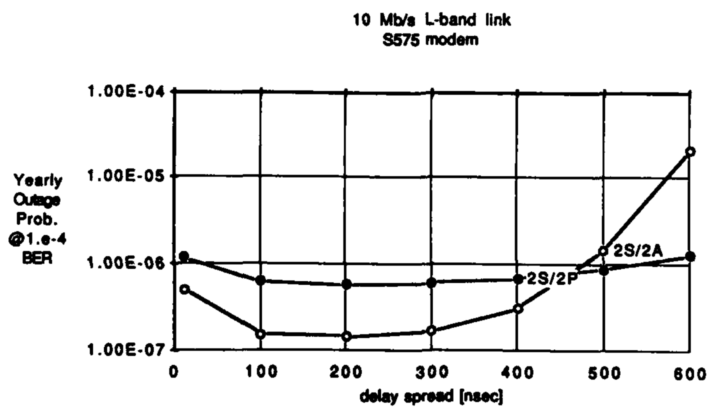


Figure 11 Lband S575 Outage

DISCUSSION

P. TWITCHELL, US

Dr. Parl is quite correct that there is a need for closer collaboration between atmospheric models and those concerned with EM propagation. Recent advances in satellite atmospheric sounders (temperature/water vapor) will provide new opportunities for EM propagation prediction.

AUTHOR'S REPLY

What is needed is a coordinated measurement program of atmospheric parameters and EM wave propagation. This would validate the known relationships between the atmosphere and radio-wave propagation. The goal would be a global propagation model that is based on global atmospheric models. The propagation model can then be constantly updated with new advances in the atmospheric sciences.

MODELLING UHF PROPAGATION FOR FREQUENCY ASSIGNMENT IN RADIO RELAY NETWORKS

Jean-Paul M.G. Linnartz[1]
Physics and Electronics Laboratory FEL-TNO
P.O.Box 96864, 2509 JG the Hague, the Netherlands.

SUMMARY

An account is given of experiences in developing UHF-propagation models for the terrain of the North German plain which is characterised by woods and farmland. A few empirical results on the relevance of modelling the interaction between diffraction and groundwave propagation are discussed. However, the paper mainly focusses on (statistical) methods used to select appropriate models, rather than going into the (physical) details of propagation mechanisms. Since local variability of the field strength prohibits prediction of the feasibility of radio relay links with absolute reliability, (static) diversity tests are recommended to avoid antenna positioning in local multipath nulls. It is shown that such antenna position tests are only worthwhile if the propagation model forecasts local mean signal powers with sufficient reliability.

1. INTRODUCTION

The effectiveness of the planning of radio links and frequency management heavily depends on the availability of appropriate propagation models. Almost from the beginning of experimental radio communications, the properties of waves travelling above the earth's surface have been a subject of research. Especially because saturation of the radio spectrum forces towards higher frequencies, the influence of obstacles and irregularities on the diffraction and scattering of waves becomes higher and more difficult to predict.
In frequency assignment for a complex radio relay network, evaluation of many path profiles is required in order to select the optimum location and frequencies for a relay node in agreement with a tactical scenario. This does not concern operational paths only because paths over which interference signals may travel have to be considered. Further, local EMC issues have to be examined. Since relay nodes are relocated frequently, relatively simple algorithms have to be used.

Forecasting path loss from terrain data can be seen as a three-step process, as illustrated in Fig.1.

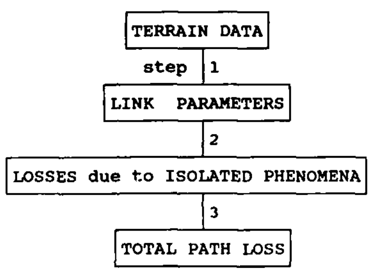


Fig.1: Three steps towards a model for propagation forecasts

The major part of the theoretical work on propagation concerns the second step only: given a set of path parameters, the loss due to isolated and idealised mechanisms can be calculated. Well-known examples are the diffraction loss A_D over an obstacle if the knife-edge geometry is known [1]-[3], or the reflection loss A_r in a two-ray model given antenna heights, and the distance of propagation over a plane, and smooth earth [4]. The reliability of the prediction of a field strength further depends on the first and third step. The first step is the interpretation of the terrain data by extracting propagation parameters. It is mainly in this step that many implementation decisions have to be made. Typical questions are "What is the effective height for diffraction of a hill covered by trees in winter?", "When does a terrain irregularity act as one single obstacle (SKE: single knife edge) and when should it be treated as two (or more) separate hills (MKE: multiple knife edge)?" and "What is the effective height of an antenna in an irregular, sloping and forested terrain?" [5],[6].

[1]Now with: Telecommunication and Traffic-Control Systems Group (TVS)
Faculty of Electrical Engineering
Delft University of Technology
P.O.Box 5031, 2600 GA DELFT, the Netherlands.

The third step describes the transition from theory to application. Experience and expertise are required to select and combine the relevant effects. The extent to which mechanisms are present (and mutually interact) is an important consideration in the selection of a model. Although usually the total path loss is estimated by linearly adding dB's as predicted in step 2, and thus assuming superimposed isolated mechanisms, this is neither theoretically justifiable nor the best solution in terms of accuracy [7],[8]. In contrast to the sequence for forecasts (Fig.1), for research and analysis, usually a reserve sequence is followed, starting with step 3.

Analysis of narrowband propagation measurements [8] in the UHF bands (bands I and II) for the "ZODIAC" 1(N1)1k radio relay network is reported in this paper. Operational links cover 5 to 30 km over the slightly hilly terrain (Δh is 20 to 100m) of the North German plain. The terrain is mainly agricultural, with scattered woodlands. Antennae heights range from 4 to 21 meters, which is often insufficient to exceed the tree-top level. Multipath reception and shadowing [4],[9],[10] introduce substantial local fluctuations of the field strength. Interaction of the radio wave with the earth's surface (reflection loss) and diffraction are of nearly equal importance in large-scale field strength fluctuations.

2. AVERAGE ATTENUATION

A typical example of modelling propagation is estimating the average loss A (in dB) with distance d, according to the statistical model

$$A = 10\beta \log d + \alpha, \qquad (1)$$

with α and β empirical constants. Theoretical values are β=2 in free space and β=4 in the case of propagation over a plane earth. An instructive conclusion on the relevance of propagation mechanisms can be drawn from experiments such as presented in Fig.2. In our application, it appeared that UHF propagation in forested areas ($\beta\approx$2.6) is principally a diffraction effect in free space, while for relatively open areas ($\beta\approx$3.8) the plane earth model is more appropriate. The effect of lower β in areas with dense forests has been investigated by Tamir [11] for VHF frequencies.

As suggested by Fig.2, adding theoretical diffraction and reflection loss (A_D + A_r), and thus β>4, can not lead to an optimum model since interaction of both mechanisms is then neglected. An appropriate and accurate VHF or UHF model is thus expected to smoothly change, e.g. as in [7], from a diffraction model for paths mainly covered by vegetation to a plane earth model for paths over farmland. Dedicated measurements near the Haarlerberg (Fig. 3) confirmed that an obstacle, though introducing diffraction loss, significantly tempers reflection loss. Field strength measurements and antenna height gain tests, gave the impression that diffraction and reflection showed their influence on different parts of the path.

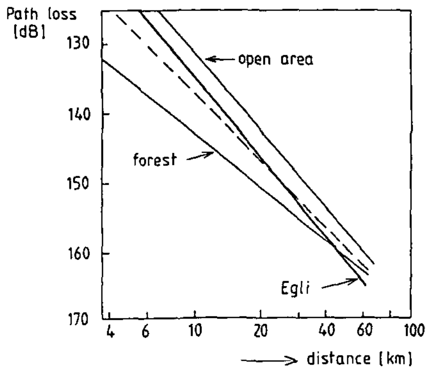
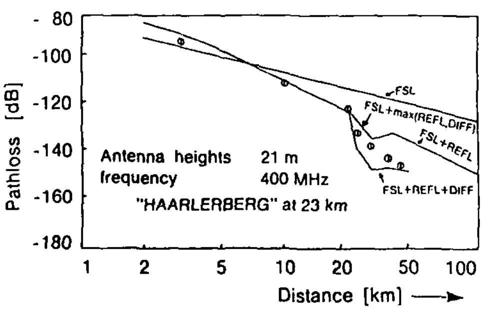


Fig.2: Average path loss in the North German plain at 300MHz



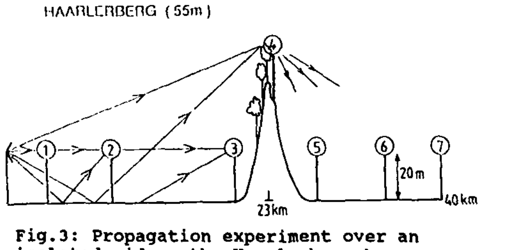


Fig.3: Propagation experiment over an isolated ridge: the Haarlerberg has a height of 55 meters above a smooth, flat ground in Holland. Propagation is best characterised by free space loss at (1), plane earth loss at (2,3), diffraction at (4) and a combination of diffraction and reflection at (6,7).

3. STANDARD DEVIATION OF PREDICTION ERROR

The logarithmic standard deviation σ_t of the error experienced when a propagation model A is verified, is defined as

$$\sigma_t^2 \;\hat{=}\; E[\,(A_m - A - E[A_m - A]\,)^2\,] \qquad (2)$$

with A_m the measured path loss (in dB) and A the forecast loss (in dB). The variance σ_t^2 is due to four statistically independent effects: model deficiencies σ_p, shadowing σ_s, multipath σ_R and measurement errors σ_m. So, if attenuation mechanisms are assumed multiplicative, σ_t is given by

$$\sigma_t^2 = \sigma_m^2 + \sigma_p^2 + \sigma_s^2 + \sigma_R^2. \qquad (3)$$

Using state-of-the-art equipment, measurement inaccuracies (though present) may be neglected ($\sigma_m \approx 1$ to 2 dB). Nevertheless, the measurements can greatly influence the results [11]. Faulty or loose connecters, damaged antennae, overloaded amplifiers etc. are not always spotted immediately and tend to introduce some outliers in the set of measurement data, rather than smoothly increasing apparent σ_t^2 by a small extra term. The selection of paths for measurements may not be limited to paths known to give satisfactory communication. Therefore, equipment has to be significantly more sensitive than operational radio relay sets, otherwise, absence of paths with heavy loss biases the population of measurements.

Shadowing and multipath reception introduce local fluctuations, usually in the range 4 to 12dB. Both effects are usually discussed in relation to mobile cellular radio-telephony [4],[9]. Assuming reasonably good antenna position, multipath reception due to local scatters gives the amplitude of the signal a Rician distribution over an area of a few tens of wavelengths. In the worst case, i.e. when no dominant wave is present as we experienced in a forest, the Rician distribution goes into a Rayleigh distribution, with $\sigma_R \approx 6$ dB (App.A). Since local phase cancelling can not be predicted, one preferably reduces multipath fluctuations σ_R before statistical processing [12],[13],[14]. Appendix A illustrates that σ_r decreases roughly with $\sqrt{m}$, where m is the number of uncorrelated samples taken at one antenna location.

Due to shadowing, the local mean power varies about the area mean with a log-normal distribution: the power in dB is normally distributed with standard deviation σ_s. In this paper, we define shadowing to occur in an area with a size equal to the resolution of the terrain data. This is in contrast to common definitions in mobile propagation, where the area is only limited in size by the requirement that shadowing remains statistically stationary. In the latter case, shadowing is typically measured over an area of a few hundreds of meters. Using the former definition, shadowing poses an insurmountable limit in achievable accuracy of forecasts ($\sigma_t > \sigma_s$). This implies that for terrain data with very high resolution, the effect of shadowing necessarily vanishes ($\sigma_s \to 0$). For average terrain, $\sigma_s \approx 12$dB when no terrain profile (except the propagation distance d) is used [15],[16].

The remaining term σ_p represents model deficiencies. The aim of the assessment of an appropriate model is to effectively reduce σ_p to zero. In (1), even if $\sigma_p \approx 0$, i.e., if β is optimally established, σ_t will always be greater than $\sigma_s \approx 12$dB. Errors $\Delta\beta$ only have a minor influence on σ_t. For samples homogeneously distributed for "log d" in the range 5 to 50 km, σ_p will equal

$$\sigma_p = 10\,\Delta\beta\,/\sqrt{12} \approx 2.9\,\Delta\beta \quad \text{(in dB)}, \qquad (4)$$

which is almost always much less than 12 dB. Nevertheless, the forest/open-area conclusion from Fig.2 appeared relevant in selecting a model (step 3). The gain from such experiments is only found in a later stage of the development of a model when effects from diffraction and reflection are carefully accounted for. Evaluation of representative profiles showed that the average Deygout [2] diffraction loss A_D is 20 dB with a standard deviation of 7 dB. If, on the other hand, a plane earth is assumed, the two-ray model yielded reflection losses A_r of about 12 dB, with a σ of 7 dB. It may be concluded that model errors σ_p are to be expected in the range of 7 dB if diffraction or reflection are not carefully considered.
In conclusion, well-known models such as [17],[18], can reach standard deviations of about $\sigma_t = \sqrt{\sigma_s^2 + \sigma_R^2} \approx 6$ dB if terrain data is carefully used. Only very complex models perform better, provided that antennae can be placed with sufficient clearance ($\sigma_R \approx 0$). The accuracy of most models, however, can be characterised by a standard deviation σ_t of 8 to 10 dB [8].

4. REQUIRED NUMBER OF MEASUREMENTS

Theory on combined occurrence of propagation mechanisms is scarce. As far as empirical results are reported, e.g. in [5]-[7], the validity in a different application is not guaranteed in advance and verification is required. Dedicated experiments, such as in Fig.3, can yield important insight into selecting and combining relevant mechanisms (step 3). Absence of local scattering and shadowing appears a relevant condition for the validity of single path experiments. However, usually a large set of samples is required to be able to draw statistical conclusions.

We assume a test model A to be verified by m samples at n different pairs of antenna locations. Multipath attenuation may be assumed statistically uncorrelated for all m measurements at one location because these are sampled at sufficiently different frequencies or antenna positions. Shadowing and model deficiencies are assumed equal (correlation unity) at one location but uncorrelated for different locations. The variance in the estimate for β can then be expressed as

$$\mathrm{var}[\beta] = \frac{\sigma_s^2 + \sigma_R^2/m}{n\ \mathrm{var}[10\log d]} \qquad (5)$$

We now define σ_{p0} as the model error experienced when the propagation factor under test is unjustly ignored. In this case, σ_{p0} is experienced when β=0 would be inserted in (1). Eq.(5) goes into

$$\mathrm{var}[\beta] = \frac{\beta^2}{\sigma_{p0}^2}\left[\frac{\sigma_s^2}{n} + \frac{\sigma_R^2}{n\ m}\right] \qquad (6)$$

Since σ_{p0} and the local variability are often in the same order of magnitude (about 7dB), the number of locations n to be investigated is in the order of 100 for a relative accuracy of 10%. Further, eq.(6) suggests that if a total of mn measurements can be taken, all measurements are preferably taken at different locations (m=1, large n). In practice, it is usually easier to take several samples at one location (m>>1) than to move to another site. Screening of data and evaluation of specific phenomena on individual paths can only be performed if multipath spreads are effectively reduced (m>>1).

In the analysis of step 1 (interpretation of the profile), the problem is more complicated. Few experiences are reported, while the number of degrees of freedom in defining effective heights and other compensation parameters appears unlimited. Experiments in step 1 are not as straightforward as linear parameter estimation because implementation decisions often have a discontinuous effect on path loss. A small change in terrain profile, for example changing the modelled effective diffraction height of trees, can have a step-wise influence of predicted diffraction loss if the implemented algorithm steps from single knife-edge to double knife-edge interpretation. Further, conclusions are sensitive to outliers. These effects are very difficult to optimise, and no general method can be recommended here. At least taking sufficient samples to smooth out discontinuities can be advised. It appeared from our measurements that 55 paths (n=55, and m=28) did not fully smooth out these effects. Practical experience indicated that the discontinuities sharply diminish with the accuracy of the implemented model. This is in contrast to the fact that a more detailed model generally has to take more decisions. The MKE diffraction model [2] gave a significantly better defined optimum than SKE methods.
Some of the parameter estimates concluded from a set of n=34 paths had to be revised after additional measurements became available (n=55, m=28). Linear regression or correlation of the error with terrain parameters, rather than studying the overall accuracy σ_t, can give some insight into the problems but careful interpretation is required.

5. LINEAR REGRESSION OF THE FORECAST ERROR WITH PATH PARAMETERS

A direct verification of decisions in step 3 can be made by correlating the residual error e ($e \triangleq A-A_m-E[A-A_m]$) with isolated losses as calculated in step 2. Linear regression in [8] showed that excess path loss (above free space A_{fs}) is proportional to 45% of the SKE diffraction loss as in [1]. It appeared that MKE models were required, since e was found proportional to 85% of the theoretical loss A_D over a Deygout triple knife edge [2]. Apart from deficiencies in implementing step 1, a very likely explanation of the lacking 15% is that propagation over smooth paths experiences relatively more reflection loss.
On the other hand, regression of the error e with two-ray reflection loss A_r was in the order of only 50% [2]. This is mainly due to the inverse frequency dependance of reflection loss, while diffraction and shadowing increase with frequency. We conclude that direct correlation of the error with losses from isolated and idealised phenomena thus prone to yield somewhat ambiguous results since losses have a mutual statistical correlation.

Correlation with distance is usually a more reliable indication of shortcomings in the calculation of reflection loss. Diffraction statistically also increases with distance, however, this effect was found to be limited to 10% of β. Correlation with frequency is a more complex parameter: diffraction, reflection and shadow losses all depend on frequency. Accumulated inaccuracies in various terms of the model are likely to require a final empirical correction with carrier frequency. In our case, average shadow loss has been modelled dependent on local terrain features by adding $A_{1,x}$, with

$$A_{l,x} = \begin{cases} 6\ \log(f/40MHz) & \text{for a low antenna (<21m) in a forest} \\ 0 & \text{in the case of high antenna clearance (open field)} \end{cases} \quad (7)$$

or any intermediate value, depending on vegetation and buildings.

Regression of e with profile parameters is also prone to diverse interpretations. For test models without diffraction, e.g. [13], terrain roughness Δh will give a positive correlation. Once diffraction is introduced in the model, the correlation often turns to negative values, since reflection loss is higher on smooth paths. Empirical corrections for terrain roughness are not easily generalised since they depend on tactical decisions in selecting antenna sites. Nonetheless, implementation decisions in step 1 are, however, best verified by correlation of e with parameters as terrain roughness, interdecile range, density of vegetation etc. For the purpose of verification of steps 1 and 3, even defining new parameters can be useful in gaining insight. Empirical correction with these new parameters does, regrettably, not lead to reliable models.

6. RESULTS FOR PROPAGATION MODEL

In our application, diffraction has been modelled by the MKE technique by Deygout. SKE methods, for instance the main hill or [3], appeared not to be essentially better than models on the form of (1). Analysing the SKE model suggested by Bullington [3], a problem of implemention was encountered for forested terrain, because the equivalent obstacle is often only determined by trees in vicinity of the antennae. Consequently, all terrain irregualities on the propagation path are then ignored.
Our algorithm takes account of up to three obstacles. The effective height of trees is empirically set to 25 meters for coniferous forest and 15 meters for other forest. The influence of terrain irregularities on ground wave propagation appeared difficult to quantify. Several attempts have been made to empirically model the reflection coefficient as a function of angle of incidence and terrain characteristics. A small improvement can be obtained by empirically modelling reflection as a function of the density of vegetation on the path, the total diffraction loss (e.g. as in [7]) or the interdecile range Δh. No such model could be scientifically justified. The increased complexity largely increased computer time, but it did not significantly increase the accuracy of the forecasts. From this point of view, reflection loss A_r was empirically reduced by 40%, which results in the generally accepted β somewhat larger than 3.2.

Since calculation time of the algorithm had to be limited, low resolution terrain data has been applied. The data base contained information on the highest point within a $1km^2$ square. However, it appeared essential to have more detailed information about the terrain height and local terrain features at the antenna site. Using the local shadow loss (7), this resulted in the model

$$A = A_{fs} + A_D + 0.6\ A_R + A_{l,tx} + A_{l,rx}. \quad (8)$$

In accordance with earlier observations, field strength forecasts are limited to about 7 or 8 dB accuracy for antennae above tree top level. For antenna heights below 20 meters in a terrain with scattered woods ($\sigma_R \approx 6dB$), σ_t could not be reduced below 9 dB.

7. NETWORK MANAGEMENT

While the propagation researcher can reduce σ_R by his measurement method (m>>1), the frequency manager has to account for a residual inaccuracy due to multipath ($\sigma_t > \sigma_R$). Local effects can not be fully predicted and accounted for but they have to be solved at the receiver and transmitter sites. If the transmission bandwidth is not too wide, a static diversity technique (e.g. testing a few alternative antenna positions) may be applied. In App.B it is shown that, although accuracy is necessarily limited, optimising σ_p (accurate model) and σ_s (high resolution) remains relevant. Conditional probabilities of successfully setting up a link suggest that only if the forecast of the local mean power has been performed sufficiently accurately and if multipath scattering is relatively severe ($\sigma_p^2 + \sigma_s^2 < \sigma_R^2$), local tests are worthwhile. Horizontal displacements of the receiving antenna showed that $\sigma_R \approx 2dB$ for antennas with high clearance. For low antennae close to woods, the standard deviation due to multipath is significantly higher ($\sigma_R \approx 6dB$). Occasional nulls of 40 dB have been recorded.
Horizontal displacement of the antenna by a few meters may be feasible with some types of masts. With other types, however, position tests might be prohibitively time consuming.
Empirically, an average height gain of 4 to 5 dB/oct was found. However, samples at intervals of 3m did not exactly follow a smooth dB/oct trend: multipath scattering produced peaks and dips with a standard deviation of 2.5dB for 250MHz, slowly increasing to 3.5dB for 900MHz. Shadowing only influences the height gain if obstructions are very close to the antenna. An exponential height gain of 1 or 2 dB per meter was measured closely beyond the Haarlerberg (Fig.3, loc.4). Within small woods, an occasional 10dB is to be gained by placing the antenna between the trunks (at 7 to 10m), rather than at the height of the leaves. In general, however, vertical

(height) tests have not been found to be very effective since the height gain prohibits all too dramatic changes (reduction) of antenna height.

Measurements indicated that the local mean power is virtually constant with time, though the multipath pattern may fluctuate somewhat. Seasonal effects [23] are present but below other statistical effects. Interference powers over longer paths (d>40km) can fluctuate due to incidental transhorizon propagation under specific conditions. It is our impression that a transportable UHF link, once in operation, experiences outage mainly due to changing interference characteristics.

8. CONCLUSIONS

To assess the optimum implementation of obtaining propagation parameters from terrain data (step 1), many measurements are required. The transition from these parameters to prediction of loss if all effects would occur isolated (step 2), has been extensively covered in literature. However, experience, expertise and dedicated measurements are required to combine these phenomena in an appropriate model (step 3).

Empirical verification of UHF-propagation models, and their implementation, is deemed necessary if an accuracy better than 10 dB is required. Computer processing time rapidly increases if higher accuracy is required. Model accuracy better than about 6 dB is usually not achievable in typical UHF military radio-relay applications because of Rician scattering of waves in the direct environment of the antenna. Our measurements indicated that the agreement between mobile propagation in cellular radio and radio relay with low, camouflaged antennas in a scattering environment is surprising. In contrast to fast fading (as for instance in civil cellular telephony), the radio relay signal barely fluctuates with time; antenna masts may thus by accident be placed permanently in local multipath nulls.

The effectiveness of repositioning the antenna is evaluated (in App.B) from theoretical models on local variability. No empirical verification of the relevance of second attempts under operational circumstances is available. Nevertheless, it was learned that reliable management of the radio relay network calls for cooperation between the frequency administration and local operators, rather than conveying a computer decision to set up a link between two sites on a prescribed frequency. Shadowing and local multipath effects cannot be fully predicted and accounted for but have to be addressed locally at the receiver and transmitter sites, e.g. by (static) diversity techniques or antenna displacement tests. The frequency manager can, however, complement these efforts by providing alternative frequencies to avoid local multipath cancelling. These conclusions may ask for modifying some of the operational procedures now used in establishing radio relay links, especially if higher (band III) frequencies are used.
It is the impression of the author that, after implementing models at present available, inaccurate modelling of co-location EMC issues, rather than link attenuation, is now the main limit to the reliability of the frequency management. Co-location interference also exhibits severe fluctuations with small antenna displacements.

References

[1] K. Bullington, Radio propagation for vehicular communications, IEEE Tr. Veh. Tech., Vol VT-26, No. 4, Nov 1974, pp.295-308.
[2] J. Deygout, Multiple knife-edge diffraction of microwaves, IEEE Trans. Ant. and Prop., Vol.AP-14 (1966) pp 480-489.
[3] J. Epstein, D.W. Peterson, An experimental study of wave propagation at 850 Mc, Proceedings of the IRE, 1953, Vol.41, pp.595-611.
[4] W.C. Jakes, Microwave Mobile Communications, John Wiley and Sons, New York, 1978.
[5] M.L. Meeks, VHF propagation over hilly, forested terrain, IEEE Trans. Ant.& Prop., AP-31(May 1983), No.3, pp. 483-489.
[6] M.L. Meeks, A propagation experiment combining reflection and diffraction, IEEE Trans. Ant. and Prop., Vol. AP-30, No. 2, March 1982, pp.318-321.
[7] A. Blomquist, L. Ladell, Prediction and calculation of transmission loss in different types of terrain, AGARD Conf. Proc., No. 144, paper 32, Electromagnetic wave propagation involving irregular surfaces and inhomogeneous media.
[8] J.P.M.G. Linnartz, The modelling of diffraction, reflection and clutter loss in UHF radio wave propagation, FEL-TNO report 1988-28, FEL-TNO, the Hague, the Netherlands.
[9] F. Hansen and F. Meno, Mobile fading - Rayleigh and log-normal superimposed, IEEE Tr. on Veh. Tech., Vol VT-26, No. 4, Nov. 1977, pp.332-335.
[10] A. Blomquist, Local ground wave field strength variation in the frequency range 30 - 1000 MHz, in Electromagnetic wave propagation, M. Desirant, J. Michiels, Eds, New York: Academic, 1960.
[11] T. Tamir, Mixed-path considerations for radio wave propagation in forest environment, AGARD Conf. Proc. No. 144-33, NATO, March 1974.
[12] IEEE Trans. Veh. Tech., Special issue on mobile propagation, Feb 1988, Vol. VT-37, No.1.
[13] W.C.Y. Lee, Mobile communications design fundamentals, Howard W. Sams & Co., Indianapolis, 1986.

[14] M.M. Peritsky, Statistical estimation of mean signal strength in a Rayleigh fading environment, IEEE Trans. Commun., Vol. COM-21, No.11, November 1973, pp.1207-1213.
[15] J.J. Egli, Radio Propagation Above 40 MC/s over Irregular Terrain, Proc. IRE, Oct 1957, pp.1383-1391.
[16] J.P. Murphy, Statistical propagation model for irregular terrain paths between transportable and mobile antennas, AGARD Conference Proceedings, No. 70, 1970, paper 49.
[17] A.G. Longley and P.L. Rice, Prediction of tropospheric radio transmission over irregular terrain, a computer method, Environmental Sience Service Administration, ESSA Tech. Rep. ERL 79-ITS 67 NTIS acc. no 676874, 1968.
[18] Y. Okumura, et al., Field strength and its variability in VHF and UHF land mobile radio service, Review of the Electr. Communic. Lab, Vol. 16, 9-10, pp 825-873.
[19] Mathematical handbook of formulas and tables, Ed. by M.R. Spiegel, Schaum's Outline Series, Mc GrawHill Book Company, New York, 1968.
[20] Handbook of Mathematical Functions. Ed. by M. Abramowitz and I.A. Stegun. New York: Dover, 1965.
[21] G.W. Lank and I.S. Reed, Average time to loss of lock for an automatic frequency control Loop with Two Fading Signals and a Related Probability Density Function, IEEE Trans. Information Theory, Vol. It-12, No.1, pp.73-75, January 1966.
[22] K. Diakoku, H. Ohdate, Optimal channel reuse in cellular land mobile radio systems, IEEE Trans. Veh. Techn., Vol VT-32, No. 32, August 1983, pp. 217-224.
[23] K. Loew, UHF measurements of seasonal field-strength variations in forests, IEEE Trans. Veh. Techn., Vol. 37, No. 3, August 1988, pp.121-124.

APPENDIX A: Standard deviation of a Rayleigh fading signal

In a radio relay network, multipath reception can be described by Rician scattering. In the worst case, the direct wave does not significantly dominate the scatters and the signal strength has a Rayleigh distribution. The pdf of the received signal power p is exponentially distributed, according to

$$f_P(p) = \frac{1}{p_1} \exp\left(- \frac{p}{p_1}\right) \qquad (A1)$$

with p_1 the local-mean power. The local-mean power μ_1, averaged over dB values, will be 2.5 dB below this linear average p_1. This is seen from

$$\mu_k = \int_0^{\infty} \{10 \log (p/1mW)\}^k \quad \exp(- p/p_1) \quad dp/p_1 \qquad (A2)$$

which is solved for k=1 using [19,(32.117)]

$$\mu_1 = \frac{10}{\ln 10}[-\tau + \ln(p/1mW)] \quad \approx \quad 10 \log(p/1mW) - 2.5 \text{ dB} \qquad (A3)$$

where τ the Euler's constant ($\tau \approx 0.577$). From [19, 32.119], the logarithmic standard deviation is found to be

$$\sigma_R^2 = \frac{100}{\ln^2 10}\left[\frac{\pi^2}{6} + \tau - \tau^2\right] \quad \approx \quad 6 \text{ dB} \qquad (A4)$$

The sum or the average of m samples has a Nakagami m-distributed envelope, or equivalently a gamma distributed power. Solutions for the higher order logarithmic moments μ_k have not been found. An approximation [13] for σ_R can be made from linear moments u and s,

$$\sigma_R^2 \quad \approx \quad 10 \log\left[\frac{u + s/2}{u - s/2}\right] \qquad (A5)$$

For Rayleigh fading, $u=p_1$ and $s=p_1^2$, so $\sigma_R \approx 4.8$ dB. For the gamma distributed sum of m samples, one finds a linear average $u=np_1$ and a second moment equal to $u_2=(n^2+n)p_1^2$. Thus, for large m, the σ_R decreases inversely proportional to $\surd m$, since

$$\sigma_{R,m} \approx 10 \log\left(1 + \frac{1}{\surd m}\right) \approx \frac{10}{\ln 10 \surd m} \approx \frac{4.3}{\surd m} \approx \frac{\sigma_R}{\surd m} \qquad (A6)$$

Appendix B: Probability of success in establishing a microwave link

In this section, benefits from applying a static diversity technique to avoid an accidentally bad position of the antenna are discussed. Two approximations are applied: the model is based on narrowband propagation, i.e., the bandwidth is assumed not to exceed a few hundreds of kHz [4]. Further, Rayleigh scattering is assumed. This represents the worst case of scattering, however, in this event, maximum diversity gain can be achieved. Results will thus tend to be optimistic for the effect of diversity.
The results presented here originally stem from (unpublished) research at D.U.T. on protocol design in mobile packet radio networks. In the original application it was concluded that the waiting time for retransmission of a lost packet can be short if shadowing fluctuates with σ_s less than 6 dB. On the other hand, if shadowing is characterised by larger σ_s, the protocol has to wait until the vehicle has moved several tens of meters to experience uncorrelated shadowing attenuation. This application comes close to common practice in Combat Net Radio (CNR), where it is known that moving the vehicle a few meters can have a dramatic influence on the success of communication.

Due to shadowing and model inaccuracies, the (experienced) local mean p_l of the received signal is log-normally spread about the predicted area mean p_a. Secondly, the actual received signal power p_s is a statistical variable with Rayleigh statistics about the local mean p_l [9]. If the power p_s is by accident below the threshold p_0, then the attempt to establish a radio link fails. Changing the antenna position gives a new (statistically uncorrelated) sample from the local process of multipath cancelling but attenuation from shadowing and inaccurate prediction remains fixed. Initially, we assume a frequency management algorithm that prescribes the use of a certain radio channel if the forecasted area-mean signal level p_a has sufficient margin α above a threshold p_0 ($\alpha = p_a/p_0$, $\alpha > 1$). In the second part of this appendix, the threshold is assumed to be dictated by interference signals rather than a known noise floor. In the latter case, the threshold is treated as a log-normal statistical variable. In both cases, the relevance of the second attempt is best evaluated from the conditional probability of success, according to

$$\text{Prob}(\text{success}_2 | \text{failure}_1) = \frac{\text{Prob}(\text{success}_2 . \text{failure}_1)}{\text{Prob}(\text{failure}_1)} \qquad \text{(B1)}$$

$$= \frac{I(1) - I(2)}{1 - I(1)} \qquad \text{(B2)}$$

with I(k) the probability that k successive attempts are all successful.

1. Noise limited network

Assuming a fixed threshold p_0, the probability I(k) is of the general form

$$I(k) = \int_0^{\infty} \exp(-k\frac{p_0}{p_l}) \frac{1}{\sqrt{2\pi}\,\sigma\, p_l} \exp(-\frac{(\ln p_l - \ln p_a)^2}{2\,\sigma 2})\, dp_l \qquad \text{(B3)}$$

which is a function of the relative margin α and the model accuracy σ in natural units ($\sigma = 0.23026\sqrt{\sigma_p^2+\sigma_s^2}$). The integral is rewritten using $\exp(\sigma t\sqrt{2}) \triangleq \alpha p_l/p_0$ to allow application of the Hermite numerical method [20, ch 25]. Thus

$$I(k) = \sum_{i=1}^{N} \frac{w_i}{\sqrt{\pi}} \exp(-\frac{k}{\alpha}\exp(\sigma x_i\sqrt{2})) \qquad \text{(B4)}$$

Here w_i are weight factors at sample points x_i for an N-point integration [20]. Figs.4a and 4b illustrate the probability of success in the first attempt I(1) and second attempt (B2). The local mean forecast error $\sqrt{\sigma_p^2+\sigma_s^2}$ is 4 and 8 dB, respectively.

For a rough model ($\sigma_t \approx 12$dB, not illustrated), the probability of success in a second attempt becomes as low as 20 to 30% even for forecast margins α as high as 10 to 20 dB. A conclusion, which is intuitively agreeable, is seen from the conditional probability of success in a second attempt (broken lines): Performing a second attempt is only worthwhile if the forecasting algorithm is accurate ($\sqrt{\sigma_s^2+\sigma_p^2} < \sigma_R < 6$dB). On the other hand, if shadowing and model inaccuracies dominate local multipath variability, the most reasonable conclusion from the failure of a first attempt is that the frequency assignment algorithm produced an erroneous forecast. A second attempt then is likely to fail again. More in general, one may conclude that improving the reliability of network management requires both improved models and operational techniques, including diversity. Improving the model can not compensate inaccuracies due to multipath. Modifying operational instructions without improving the propagation model will only frustrate local operators.

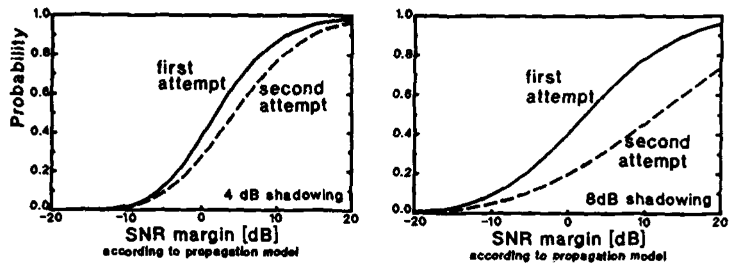


Fig.4: Probability of immediate success in establishing a radio link in a Rayleigh scattering environment, and the conditional probability that a second attempt is successful given the failure of a first attempt. Combined shadowing and model deficiencies are 4 and 8 dB, in Fig.4a and 4b, resp.

2. Interference limited network

The concept of [22] is applied for the event of one dominant interferer. The frequency assignment algorithm prescribes the use of a frequency if the signal power is expected to exceed the interference power by at least the threshold ratio z. The forecast margin above this threshold is denoted α. The desired signal and accumulated interference have correlated shadowing. For low antennae in a forest, the correlation ρ was found in the range 0.1 to 0.4, while $\rho \approx 0$ in an open field. Since the model accuracy σ_p was found relatively insensitive to the propagation distance, the effective logarithmic standard deviation σ_e is found from

$$\sigma_e^2 = (0.23026)^2 \quad (\sigma_p^2 + \sigma_s^2)\ 2\ (1-\rho) \tag{B5}$$

The ratio w of the local-mean powers of signal and interference is log-normally distributed about the (forecast) area mean ratio αz

$$f_W(w) = \frac{1}{\sqrt{2\pi}\ \sigma_e\ w} \exp\left(- \frac{\ln^2 w/\alpha z}{2\ \sigma_e^2}\right) \tag{B6}$$

The probability I(k) of k successive successes is known from diversity techniques in cellular radio telephony [22,(8)]

$$I(k) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left[\frac{\alpha}{\alpha + \exp(2\ \sigma_e\ t)}\right]^k \exp(-t^2)\ dt \tag{B7}$$

Using (B2), the conditional probability of success in a second attempt is calculated and portrayed in Fig.5.

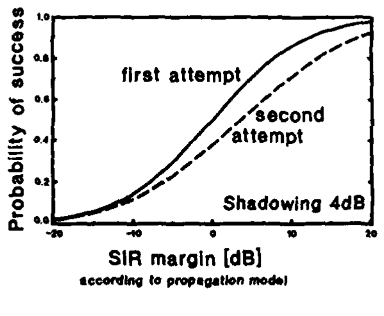


Fig.5: Probability of success in establishing a radio link in an interference limited network with Rayleigh scattering channels. Combined shadowing and model deficiencies is 4dB.

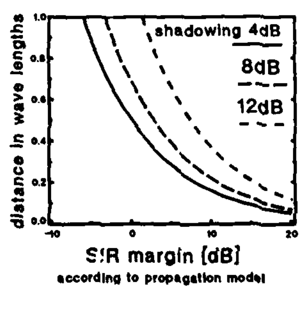


Fig.6: Expected required horizontal displacement of the antenna position as a function of the forecast margin for models of various accuracy.

Assuming exponentially distributed fade lengths, the average distance over which the C/I-ratio drops below the threshold z equals the average distance δ required to move an antenna out of a multipath null. Given the local mean C/I ratio w, δ is found from rewriting the average fade duration for two contending Rayleigh distributed signals [21].

$$\frac{\delta}{l} = \frac{\sqrt{2}\,\sqrt{z}}{\pi\,\sqrt{w}} \tag{B8}$$

Averaging over the log-normal distribution (B6), one finds

$$\frac{E[\delta]}{l} = \frac{\sqrt{2}\,\sqrt{z}}{\pi} \int_{w=0}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma_e\,\sqrt{w}} \exp\left(\frac{-(\ln w - \ln \alpha z)^2}{2\,\sigma_e^2}\right) d\ln w \tag{B9}$$

Substituting $w = \alpha z \exp(\sqrt{2}\,\sigma_e x)$ results in

$$\frac{E[\delta]}{l} = \int_{-\infty}^{\infty} \frac{2\sqrt{z}\,\exp(-x^2 - \frac{1}{4}\sqrt{2}\sigma\,x)}{\pi\sqrt{\alpha z}\,\sqrt{2\pi}\,\sigma}\,dx = \frac{2}{\pi\sqrt{\alpha}\,\sqrt{2}} \exp\left(\frac{\sigma^2}{8}\right) \tag{B10}$$

The expected displacement of the antenna as a function of the predicted margin α is shown in Fig.6. The curves suggest that if accidental bad positionig of the antenna is to be resolved by position tests, only minor displacements are required, since the wave length is typically less than one meter.

DISCUSSION

D. YAVUZ, TUR

Was the US-ECAC model TIREM (Terrain Integrated Rough Earth Model) which uses digitized terrain data plus multiple diffraction, reflection and other sub-models considered for your application?

AUTHOR'S REPLY

The analysis was initalized with relatively simple models because of our special application for rapid frequency assignment with relatively limited computer support. Further, we felt that converting a propagation model to a new database is not always trivial. Nevertheless, TIREM can be of interest as soon as the computer used for frequency assignment is upgraded. In present investigations on propagation in Band III (1300-1800 MHz), TIREM is one of the models considered in the analysis at FEL-TNO.

PATTERN RECOGNITION TECHNIQUES APPLIED TO THE NASA-ACTS ORDER-WIRE PROBLEM

Kenneth C. Allen
U.S. Department of Commerce
National Telecommunications and Information Administration
Institute for Telecommunications Sciences
325 Broadway
Boulder, CO 80303-3328
USA

SUMMARY

The National Aeronautics and Space Administration (NASA) is planning to launch an Advanced Communication Technology Satellite (ACTS). This satellite will use several methods to adaptively mitigate the effects of fading due to rain in its 20- and 30-GHz links. A decision algorithm is needed to control the implementation of these mitigating techniques. The use of a learning pattern recognition algorithm as the solution to this problem is explored here. The algorithm is applied to data simulating fading on a slant path through the atmosphere at 20 and 28.8 GHz. The data were scaled from data measured on a 1-km path at 28.8 GHz in Hilo, Hawaii. The sole cause of fading in the data was attenuation due to rainfall on the path. The results indicate, at least for the test data, that the use of the most recently measured value of the received signal level (RSL) is adequate for predicting the future RSL and can be used to decide when the mitigating techniques should be used. More complicated algorithms based on linear combinations of previous RSL samples do not perform significantly better than the simple previous value approach.

1. INTRODUCTION

The National Aeronautics and Space Administration (NASA) is planning to launch an Advanced Communication Technology Satellite (ACTS) in 1992. The satellite and the ground stations will have the ability to adaptively increase the transmit power in the 20-GHz down link and 30-GHz up links to overcome fading due to rainfall on the path. An algorithm is needed to decide when the power should be boosted.

In the study reported here, learning pattern recognition techniques are used to develop and explore the performance of various algorithms. A pattern recognition technique was chosen because of its flexibility. Although, only the received signal level (RSL) for past time is used as the information input to the algorithm, many other inputs can be combined in a pattern recognition algorithm, including radar data, weather forecasts, rain rate gauge measurements, etc.

Data, simulating the fading to be experienced on a 20-GHz down link and a 28.8-GHz up link, are generated by scaling data measured at 28.8 GHz on a 1-km link in Hilo, Hawaii. The data are then separated into a training set and a testing set. The training set is used to teach the decision algorithm to make good decisions concerning when to boost the transmit power. The testing set of data is used to test the performance of the decision algorithms and determine their relative performance.

The results of the testing reveal the characteristics of the fading in the simulated data and indicate that the single most recent value of the RSL is in many ways the best near-term predictor of the future RSL. Thus, in actual operational practice after ACTS is launched, the following approach to implementing a decision algorithm is indicated. First, the most recent value of the RSL should be used to decide whether or not transmit power should be boosted (last point algorithm). While this approach is being used, a learning decision algorithm should be trained on the RSL data observed at each ground station. After training, an algorithm based solely on past values of the RSL may only marginally outperform the last point algorithm.

2. BACKGROUND

The ACTS satellite is scheduled to be launched in 1992. The purpose of ACTS is to test and develop advanced satellite communication technology. The satellite will operate in the 20/30-GHz band with spot beams that can be rapidly switched among ground stations. The spot beams will dwell at each location for a few milliseconds and return to the same

location at least once a second. The switching and dwell times are adaptively controlled in response to the communication traffic load by the satellite control center.

A low burst rate (LBR) link will incorporate processing of the data on board the satellite. Data received from a ground station are demodulated and stored. When the spot beam moves to a new location, all the data that has accumulated on board the satellite and are addressed to the ground stations illuminated by the spot beam are transmitted. The satellite will also have a high burst rate (HBR) link. The HBR link operates as a simple "bent pipe" (repeater) and spot beams are used to connect two ground stations simultaneously.

Each ground station communicates with the satellite control center through the satellite on what is called the order wire, which is provided by reserved bits in the communication protocol. The order wire enables the ground station to inform the control center of traffic requirements and the control center to inform the ground station of the next time it will receive the spot beam.

The satellite transmits continuous wave beacons near 20 and 30 GHz, which the ground stations monitor to measure the attenuation that will be experienced on the wide-band communication links. This information is then used by coordinating through the order wire to adaptively compensate for channel fading. The LBR link will use forward error correcting (FEC) coding to compensate for severe fading and the HBR link will use variable transmit power. The need for a decision algorithm to adaptively control these mitigation techniques is called the "order-wire problem." The application of a learning pattern recognition algorithm to this problem was explored and is reported below.

It is desired that an availability of 0.99999 of the time be achieved for the links.

3. ANALYSIS

In the work described below, the use of a learning pattern recognition algorithm serve as a decision algorithm for controlling the power boosts for the HBR link is explored. First, the mathematical foundation of the algorithm is reviewed. Then data simulating fading on a slant path to ACTS are simulated from data measured on a short terrestrial path. These data are separated into two sets. The first data set serves to train the learning algorithm. The second set serves to test how well the trained algorithm performs.

3.1 Learning Pattern Recognition Algorithms

A general form of the pattern recognition problem is shown in Figure 1. The observation space contains the information that is available for determining the classification of a pattern, i.e. determining to which class or group it belongs . In the case here, it includes the recorded time series of the RSL for past time. It could, in general, include radar data, rain gauge data, time of year data, etc.

The feature extraction process results in a pattern vector, x. Each element of the pattern vector may be a complicated function of parameters or raw values taken from the observation space.

In a linear decision algorithm, the dot product of the pattern vector and a weight vector, w, is computed. When there are more than two classes, each class has its own weight vector; the pattern vector is classified as belonging to the class whose weight vector results in the largest dot product. In the two-class case, the dot product is compared to a threshold to determine classification. For the work here, a two-class algorithm is sufficient. The two classes, for example, may be "the power needs to be increased" or "the power does not need to be increased" for the next transmission.

The feature extraction process is the most critical step in pattern recognition. It provides the information on which a classification decision will be made. It is especially important that the information in the pattern vector be suitable for processing by a linear decision algorithm where only linear combinations of the vector elements are allowed.

The linear decision algorithm is straightforward, and learning algorithms have been developed that enable the weight vector, w, to be determined from a training set of pattern vectors, x(k). The algorithms derived here are based on the Robins-Monro (R-M) algorithm [1], and the development and notation generally follow Tou and Gonzalez [2].

The R-M algorithm finds the root w_o of the deterministic function $g(w)$ when only noisy observations,

$$h(w) = g(w) + n, \tag{1}$$

are available. The noise must have zero mean so that

$$g(w) = E\{h(w)\}. \tag{2}$$

Then, if the variance of n is bounded for all w, the root w_o can be found from an arbitrary initial $w(1)$ using

$$w(k+1) = w(k) - a_k h[w(k)]. \tag{3}$$

The sequence $\{a_k\}$ must consist of positive numbers satisfying

$$\lim a_k = 0 \text{ as } k\to\infty, \tag{4}$$

$$\Sigma a_k = \infty, \text{ and} \tag{5}$$

$$\Sigma a_k{}^2 < \infty. \tag{6}$$

The harmonic sequence $a_k=1/k$ satisfies the above conditions and will be assumed in what follows. The regression function $g(w)$ must have only one root and must be bounded in the interval of interest. If all the above conditions are satisfied, the sequence w_k converges to the root w_o in the mean-square sense and with probability one.

In what follows, the regression function is used as the derivative of a criterion function so that finding the root of the derivative yields the minimum of the criterion function. By selecting certain criterion functions, it is possible to develop learning algorithms for deriving the weight vector for a linear decision function.

Let the criterion function be

$$J(w,x) = E\{f(w,x)\}, \tag{7}$$

where $f(w,x)$ is a random function, by which we mean that at every point (w,x), f is a random variable. The minimum of $J(w,x)$ can then be found by finding the root of its derivative. Let

$$g(w) = \delta E\{f(w,x)\}/\delta w = E\{\delta f(w,x)/\delta w\} = E\{h(w)\}. \tag{8}$$

Then substituting into (3), the w_o that yields the minimum of $E\{f(w,x)\}$ is found using

$$w(k+1) = w(k) - a_k[\delta f(w,x)/\delta w]_{w=w(k)}. \tag{9}$$

One possible criterion function is formed by letting

$$f(w,x) = |r(x)-w^T x|, \tag{10}$$

where $r(x)$ is a random variable to be defined later. Then, by substituting

$$h(w) = \delta f(w,x)/\delta w = -x\,\mathrm{sgn}[r(x)-w^T x] \tag{11}$$

into (9), the minimum of $J(w,x)$ is found using

$$w(k+1) = w(k) + a_k x(k)\,\mathrm{sgn}\{r[x(k)]-w^T(k)x(k)\}. \tag{12}$$

The sequence $\{w(k)\}$ found using (12) converges to the w_o that gives the best estimate of $r(x)$ by the product $w^T x$ in the sense that $E\{|r(x)-w^T x|\}$ is minimum. Stochastic iterative approximations such as (12) are generally slow to converge. One suggested method for increasing the speed of convergence is to increment a_k only when $r[x(k)]-w^T(k)x(k)$ changes sign.

Another possible criterion function is formed by letting

$$f(w,x) = [r(x)-w^T x]^2/2. \tag{13}$$

Then, by substituting

$$h(\mathbf{w}) = \delta f(\mathbf{w},\mathbf{x})/\delta \mathbf{w} = -\mathbf{x}[r(\mathbf{x})-\mathbf{w}^T\mathbf{x}] \qquad (14)$$

into (9), the minimum of $J(\mathbf{w},\mathbf{x})$ is found using

$$\mathbf{w}(k+1) = \mathbf{w}(k) + a_k \mathbf{x}(k)\{r[\mathbf{x}(k)]-\mathbf{w}^T(k)\mathbf{x}(k)\}. \qquad (15)$$

The sequence $\{\mathbf{w}(k)\}$ found using (15) converges to the $\mathbf{w}_o$ that gives the best estimate of $r(\mathbf{x})$ by the product $\mathbf{w}^T\mathbf{x}$ in the sense that $E\{[r(\mathbf{x})-\mathbf{w}^T\mathbf{x}]^2\}$ is minimum, i.e., the least squares estimate.

The random variable, $r(\mathbf{x})$, is called the random classification variable. The definition chosen for $r(\mathbf{x})$ depends on the objectives one hopes to accomplish by using (12) or (15). In the application to the ACTS order-wire problem to follow, $r(\mathbf{x})$ will be defined in two ways.

3.2 Training and Testing Data

Fading data are required to test the learning algorithm approach. The data available to the author were measured on a 1-km, line-of-sight path in Hilo, Hawaii. The data consisted of RSL at 28.8 GHz and rain rate recorded once each second from March through September 1988 (7 months).

These data were used to simulate fading due to rain on a ground-to-satellite path at 20 and 28.8 GHz. Rain's domain in the atmosphere is approximately from the freezing layer (zero degree isotherm) height to the ground. A typical value for the zero degree isotherm height is 4 km. A slant path to a 4-km height with an elevation angle of 67 degrees would have a projection on the ground of 1-km length. The slant range of the path would be about 4 km. Fading at 28.8 GHz due to rain on this hypothetical slant path can be approximated by multiplying the fades measured on the 1-km path by four. Assuming that attenuation by rain is approximately proportional to the square of the radio frequency, fading at 20 GHz on this hypothetical slant path can be approximated by multiplying the measured fades by two.

The simulated data were broken into two parts, a training set and a testing set. The training set was taken from the first 3 months of measured data and consisted of only those hours in which the signal faded by at least 3 dB. This data set contained 871,951 seconds, 39 percent with rainfall occurring and an accumulated rainfall of 551 mm. The testing set was taken from the 4 months and consisted only of those hours in which rainfall was measured. This data set contained 2,625,307 seconds, 25 percent with rainfall occurring and an accumulated rainfall of 845 mm. Each hour of data was normalized so that the average, unattenuated RSL was at 0 dB.

At any one second, the current and past values of the RSL can be used to make decisions about future values of the RSL. The notation to be used for these signal levels is as follows. At the current time (decision time), the RSL is X_0. The RSL measured n seconds before the decision time is X_{-n} and the RSL that will occur n seconds in the future is X_n.

4. RESULTS

Several algorithms for deciding when to boost power on the 20- and 30-GHz links were tested using the simulated slant path fading data. For the 20-GHz down link, one trivial and two intuitive algorithms as well as several pattern recognition algorithms were tested. The best performing intuitive algorithm and best performing pattern recognition algorithm for the 20-GHz down-link problem were also tested for the 30-GHz up-link problem.

4.1 20-GHz Down Link Problem

It is planned that the satellite will be able to boost power by about 10 dB for the duration of the dwell time of the spot beam to overcome fading at a particular ground station. The boost in power is planned to be used when the signal fades by 3 or more dB. The problem can be formally stated as the need to decide whether or not the RSL 1 second in the future will be below -3 dB. In the test data set, this condition exists about 1 percent of the time.

Zero-point prediction

One approach to the decision algorithm would be to simply randomly boost power for some fraction of the time. No information about the signal level would be used. This will be called the zero-point prediction because no data points are used. The results of applying this approach to the test data are shown in Figure 2. A miss is a 1-second data point for which the algorithm should have decided to boost power but did not. A false alarm is a 1-second data point for which the algorithm should have decided not to boost power but did. These are the two types of errors that the algorithm can make. If the power is never boosted, then there are no false alarms but the probability of a miss is about 1 percent for the test data. If the power is always boosted, then there are no misses but the probability of a false alarm is about 99 percent. By choosing to randomly boost power by some fraction of time ranging from zero to one, various combinations of probabilities for false alarms and misses between the above two extremes can be achieved, as shown by the curve in Figure 2.

One-point prediction

Another approach to the decision algorithm would be to simply take the value of the last 1-second sample and compare it to some threshold, e.g., -3 dB, in order to decide whether to boost power for the next data point. This algorithm can be written as

$$-X_0-3 > t. \tag{16}$$

If (16) is true, then the power should be boosted. The threshold, t, can be varied from zero to achieve different combinations of false alarms and misses. Figure 2 shows the performance of this algorithm with various values of t. For the test data, the algorithm results in many fewer false alarms and misses than the random-boost algorithm. For the best threshold value of t=0, the total error rate (number of false alarms, 660, plus number of misses, 675, divided by the total number of test points, 2,623,023) was .000509.

This algorithm serves as a standard to which to compare the performance of other algorithms. It is based solely on the persistence in the data, i.e., the tendency of the RSL to remain the same from second to second. One would like to believe that it is possible to develop an algorithm that can outperform simple persistence.

Two-point extrapolation

Yet another approach to the decision algorithm is to take the current RSL and the one before it and, using straight-line extrapolation, predict the RSL 1 second in the future. This prediction can then be compared to the -3 dB threshold. The prediction of the RSL of n seconds in the future is given by

$$\hat{X}_n = (X_0-X_{-1})n + X_0 \tag{17}$$

so that the 1-second algorithm can be written as

$$-\hat{X}_1-3 = 2X_0 - X_{-1} - 3 > t. \tag{18}$$

If (18) is true, then the power should be boosted. The threshold is again variable to create a tradeoff between false alarms and misses as shown in Figure 2.

It should be noted that the two-point extrapolation algorithm does not perform as well as the last-point algorithm (one-point prediction). For the best threshold, t=0, the total error rate (1095 false alarms and 984 misses) was .000793. One would hope that the additional information added by the second point would improve the prediction. The fact that the extrapolated prediction is worse than the last-point prediction can be explained, however, as the result of noise in the data. This noise results in the slope of the extrapolation line sometimes deviating to extremes above or below appropriate values, amplifying the noise in the prediction, so that the average prediction error is actually increased.

Two-point learning algorithm

In the classical approach to pattern classifiers for the two-class case, the value of the random classification variable, r(x), is determined by the class membership of x. In the case presented here, the augmented pattern vector, x, is $(X_0, X_{-1}, 1)$, and the value of r(x) is 1 if X_1 is less than -3 dB and -1 if X_1 is greater than -3 dB. Thus, r(x) is 1

if the power should be boosted and -1 if the power should not be boosted. The algorithm given by (15) then converges to the weight vector, $w = (W_0,W_1,W_2)$, which results in the dot product of w and x being the best estimate of r(x) in the least-squares-error sense. The decision algorithm is then based on comparing the dot product of w and x to zero. If the dot product is greater than zero, then the power should be boosted. If it is less than or equal to zero, the power should not be boosted. By again introducing a variable decision threshold, the decision algorithm is given by

$$w^T x > t. \tag{19}$$

The algorithm given by (15) was applied to the training data, and after more than 20 million iterations the resultant weight vector was (-.1057,-.1051,-1.003). The performance of (19), when applying this weight vector to the test data, was slightly worse than the last point algorithm given by (16). The results are <u>not</u> shown in Figure 2.

The r(x) selected according to the classical approach is not well matched to the dot product used to estimate it. The RSL is nearly zero for most of the time that r = -1 and varies widely from around -3 to -10 dB and lower when r=1. It is difficult for the algorithm to estimate r(x)=1 by $w^T x$ when the signal level varies over such a wide range.

A better match between the random classification variable and the pattern vector could be achieved by selecting a different pattern vector in the feature extraction process or by selecting a different random classification variable. The later alternative was explored.

The random classification variable was given the values -1 for no power boost and 10 for the power boost case. After running the algorithm given by (15) on the training data for more than 20 million iterations, the resultant weight vector was (-.5969,-.5921,-1.188). When using this weight vector, the algorithm given by (19) performed slightly better than it did for the weights found for the r=-1,1 case.

An even better random classification variable is defined by $r(x)=X_1$. In this case, the dot product of w and x is used to estimate the next value of the RSL. This is a good match between the random classification variable and the pattern vector and ultimately resulted in the best two-point prediction algorithm. It also has the advantage that if the threshold at which the power is to be boosted is changed, the same weight vector can continue to be used, while for the previous classification variables, the training algorithm would have to be reapplied to find the new w.

The decision algorithm for this latter classification variable consists of comparing the dot product estimate of X_1 to the power boost threshold of -3 dB. In its more general form, the decision algorithm is

$$-\hat{X}_1-3 = -w^T x-3 > t. \tag{20}$$

The training algorithm given by (15) in this case resulted, after about 20 million iterations, in the weight vector (.5061,.4933,-.0008). When the decision algorithm given by (20), using these weights, was applied to the test data, the resultant performance was marginally better than the last-point prediction algorithm as can be seen in Figure 2. For the best threshold, t=0, the total error rate (581 false alarms and 707 misses) was .000492.

It should be noted that the RSL is very tightly distributed about the -3 dB threshold for all the false alarms and misses. If fact, if only misses with the RSL less than -3.3 dB are considered as errors (result in an outage), then the error rate was only 0.000004. Thus, the desired availability of 0.99999 can easily be attained by the decision algorithm if a small margin is provided between the decision threshold and the outage threshold. This result, of course, depends on the transmit power boost being large enough.

It is logically required that the optimum weights for a two-point algorithm produce equal or better predictions in a least-squares-error sense than a one-point algorithm. When the performance of the one-point and two-point algorithm are compared for the training data set, it is found that the two-point case (1414 total errors) does indeed perform slightly better than the one-point case (1526 total errors) by about 7 percent. However, for the test data set, the performance increases by only about 3.5 percent. This decrease in the performance is caused by differences in the training and test data sets. Therefore, although the weights are optimum for the training set, their performance is somewhat suboptimum for the test data set. This means that the weights of the two-point prediction

algorithm must be very finely tuned to the data base in order to significantly outperform the one-point predictions. Thus, in practice, the two-point predictions are only marginally better than the one-point predictions for data not used for the training.

Ten-point learning algorithm

The training algorithm given in (15) was used to find the weights needed to predict the RSL by estimating $r(x)=X_1$ using the previous 10 values of the RSL. The augmented pattern vector was $(X_0, X_{-1}, \ldots, X_{-9}, 1)$. The resultant weights after more than 20 million iterations were (.1544,.1377,.1242,.1126,.1017,.0912,.0817,.0730,.0650,.0570,-.0010).

When the decision algorithm (20) was applied using these weights, the results were only slightly better than those of the two-point extrapolation algorithm. Again, it would be expected that better predictions could be achieved using 10 data points than using only 1 data point. However, it seemed that the learning algorithm was converging much slower when 10 weights were being found than when only 2 weights were needed. After more than 67 million iterations on the training data set, the weights were (.1615,.1429,.1278,.1143,.1021,.0902,.0794,.0694,.0603,.0512,-.0007). When these weights were used in the decision algorithm, the results were better than when the weights found after only 20 million iterations were used, but the results were still significantly worse than for the one-point algorithm. The convergence may be so slow for the 10-point case as to be impractical.

4.2 30 GHz Up Link Problem

It is planned that the ground stations will be able to boost power in steps by up to 13 dB to overcome fading. For the purposes of the study reported here, a step size of 1 dB will be used. The power boost would be used when the RSL fades by 5 dB or more. The problem can be formally stated as the need to decide whether or not the RSL 1 second in the future will be below -5 dB and if it is, then the transmit power should be increased by the appropriate number of whole decibel steps to keep the signal level above -5 dB.

In Figure 3, the sample cumulative distribution of the test data set at 28.8 GHz is shown. Also shown in the figure is the cumulative distribution when the last-point prediction of RSL is used to determine the amount of ground station transmit power increase. The decision algorithm gives the amount of power boost, P_B, as

$$P_B = \begin{cases} 0, \text{ if } X_0 \text{ is greater than -5 dB.} \\ \text{integer part of } (-4-X_0), \text{if } X_0 \text{ is less than or equal to -5 dB.} \end{cases} \quad (21)$$

As can be seen in the figure, determining the power boost from the most recent value of the RSL works well and decreases the amount of time that the RSL is below -5 dB by a factor of five to about 0.004, which is the fraction of time that the distribution without any power boost is below -13 dB. Because of the large amount of rainfall included in the test data set, about one order of magnitude less outage time would be expected for a typical climate, so that if the fade margin is such that an outage occurs when the RSL is less than -10 dB (10 dB below median signal level), the outage time would be about 0.00015 giving a reliability of 0.99995. This is almost the level of performance desired. The reliability does not seem to be limited by the performance of the power boost decision algorithm but by the amount of power boost that is available.

The amount of power boost given by the learning algorithm predicted RSL is

$$P_B = \begin{cases} 0, \text{ if } w^T x \text{ is greater than -5 dB.} \\ \text{integer part of } (-4-w^T x), \text{ if } X_0 \text{ is less than or equal to -5 dB.} \end{cases} \quad (22)$$

The resultant cumulative distribution of RSL when this algorithm is used with a two-point pattern vector is not shown in Figure 3 because it is so similar to that of the last-point algorithm given in (21).

5. CONCLUSIONS

The results presented above show, at least for the particular test data used, that the single last value of received signal level (RSL) is nearly the best predictor of the next value of RSL, at least for the restricted class of algorithms based on the prediction of RSL from linear combinations of past RSLs.

At first thought, these results seem to indicate that the RSL can be conceptually thought of as being composed of two components. One component is caused by attenuation due to rainfall and its value changes so slowly compared to the 1-second sample rate that the most recent sample very closely approximates the next sample's value and therefore additional past samples contain very little additional information. The second component is noise-like and cannot be predicted by any number of past samples. Thus, the last-point prediction (decision) algorithm is nearly optimum. The sample of test data shown in Figure 4 visually supports this conceptualization of the data.

Perhaps more accurately, however, the RSL can be thought of as composed of three parts, as given by

$$X_{rsl} = X_{rain} + X_{scintillation} + X_{noise}. \quad (23)$$

The part caused by rain is slowly varying and is 0 dB most of the time, but sometimes can decrease up to tens of decibels. The scintillation part caused by turbulence varies about 0 dB by up to a few decibels. The noise contribution varies fractions of a decibel about 0 dB.

When the data used in this analysis were scaled from the 1-km path to simulate a slant path, the multiplication left the relative amplitude of the scintillation and noise components unchanged. However, on a real slant path, the scintillation component will be determined by the longer path while the noise will be determined by the radio equipment and may be even less than in the 1-km path data. Thus, significantly different relative amplitudes will probably exist on operational slant paths. This is important because, while the noise component is not predictable, the scintillation component may be partially predictable so that a two-point prediction may perform better than a one-point prediction.

When an optimum prediction is subtracted from the data, the spectrum of the resultant difference is white, i.e., flat. If it is nonwhite, there is still information left in the data that can be removed by prediction. For many years, first differencing has been used to "whiten" scintillation data, i.e., by subtracting the previous value from the next, the spectrum is made nearly white. Therefore, even with larger, relative scintillation amplitudes, a one-point prediction algorithm may be close to the best. A two-point algorithm may always be only marginally better, even after a long training period. A nonlinear feature extraction may be required before a one-point prediction can be significantly improved upon.

Studies into the temporal statistical properties of fading have been conducted by R.M. Manning of NASA's Lewis Research Center (private correspondence, April 14, 1989). Those studies, and an extension of those presented here, should be continued in an effort to find a method better than simple persistence. However, it is likely when ACTS is launched that last-point algorithms will be the best that can be used until possibly better ones are developed based on the experience at each ground station.

The results presented indicate that a one-point algorithm should be sufficient to provide the desired availability of 0.99999 if the available power boost is sufficiently large and a small margin is provided between the decision threshold and the outage threshold for RSL.

6. REFERENCES

[1] Robins, H., and S. Monro, "A stochastic approximation method," Annals Math. Statist., vol. 22, 1951, pp.400-407

[2] Tou, J.T., and R.C. Gonzalez, "Pattern Recognition Principles," Reading, MA, Addison-Wesley Publishing Company, 1974, pp. 218-234

7. ACKNOWLEDGMENTS

The author would like to thank Howard Tatum and Steve Butler of the Weather Service Office of the National Oceanic and Atmospheric Administration in Hilo, Hawaii, for all of their assistance in the acquisition of the measured fading data at 28.8 GHz on the 1-km path. in Hilo. The measurements could not have been performed without their assistance, which was given with such good will.

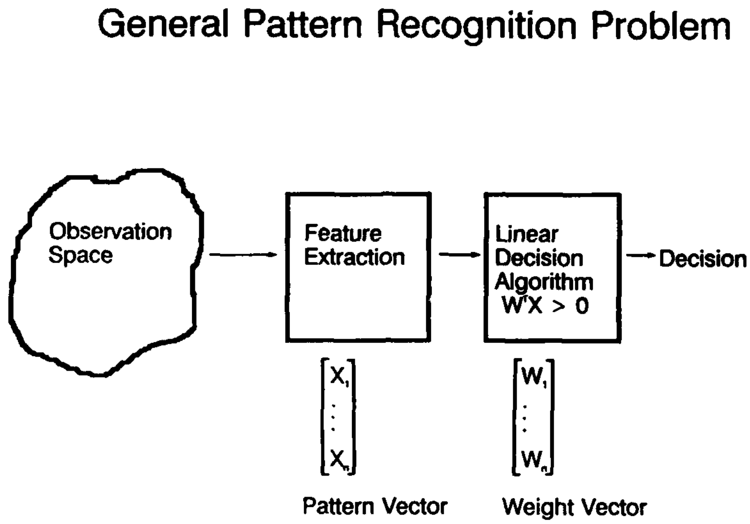


Figure 1. A diagram of the general pattern recognition problem.

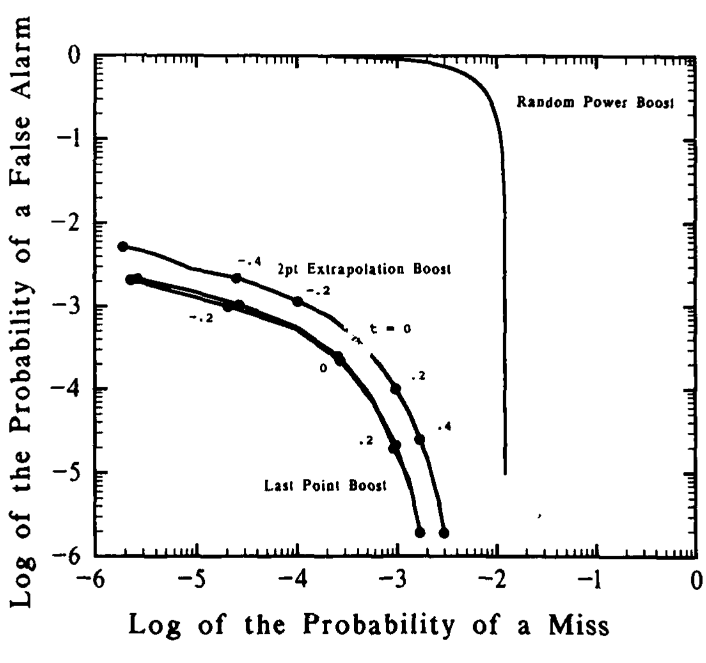


Figure 2. The performance of different decision algorithms as depicted by the probability of false alarms and misses when applied to the 20 GHz test data.

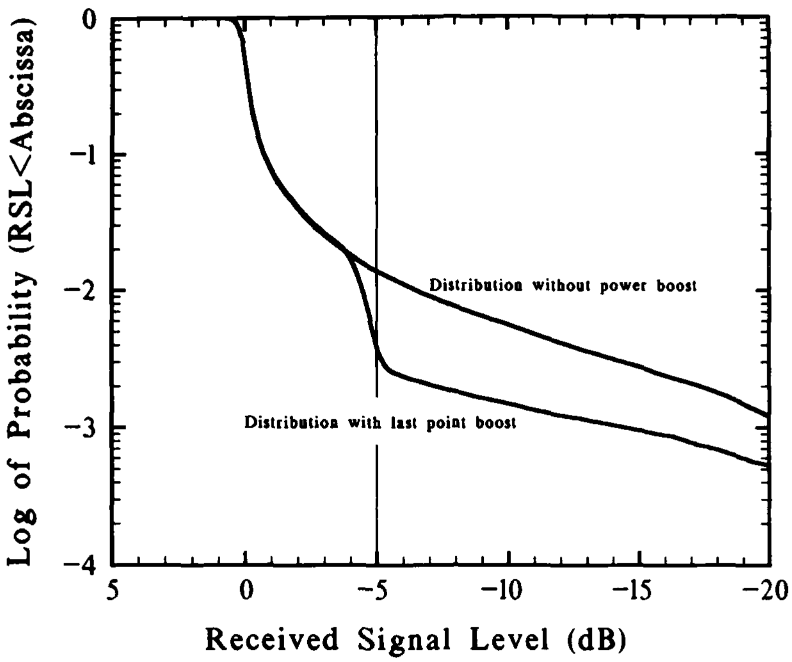


Figure 3. The cumulative distributions for the 28.8 GHz test data without power boosting and with power boosting based on the last value of the RSL.

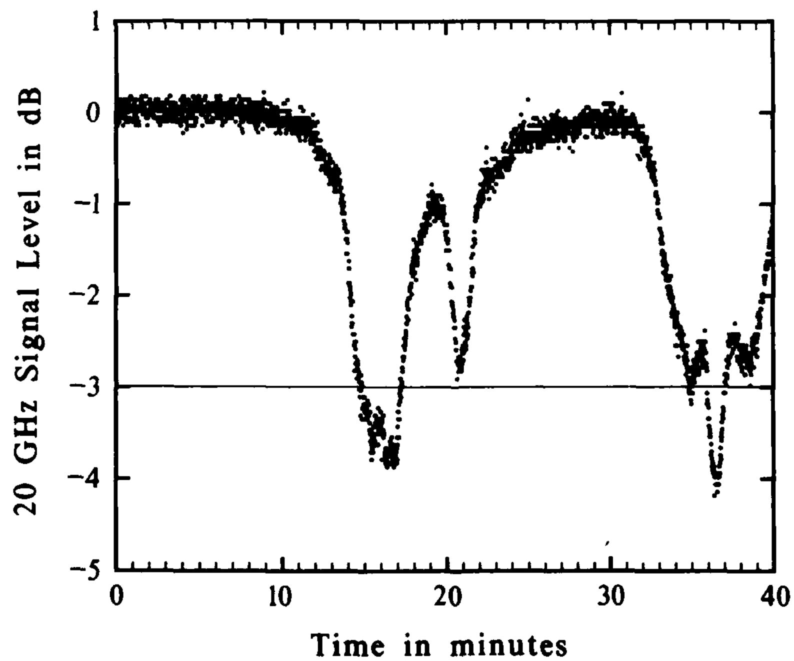


Figure 4. A sample of the 20 GHz test data.

DISCUSSION

D. YAVUZ, TUR

A linear metric is suitable for data whose underlying statistical behavior can be modeled by a first order Markov process. Perhaps a different metric and/or maximum entropy type estimation algorithm might give better results.

AUTHOR'S REPLY

The author has not examined whether or not the data can be modeled by a first order Markov process. It is an important question. The advantage of using the pattern recognition approach is that other data such as rain rate gauge and radar measurements can be easily incorporated into the decision algorithm.

The FTZ HF propagation model for use on small computers and its accuracy

by

Th. Damboldt and P. Suessmann
Forschungsinstitut der DBP beim FTZ
Postfach 5000
D-6100 Darmstadt

Abstract

In the first part of the paper (sections 1 to 6) a self-contained method of estimating the critical frequency and the height of the ionosphere is described. This method was implemented in the computer program FTZMUF2. The accuracy of the method tested against the CCIR-Atlas (Report 340) yielded an average difference of less than 0.1 MHz and a standard deviation of 2.3 MHz.

In the second part of the paper (sections 7 to 13) the FTZ HF field-strength prediction method is described which is based on the systematics found in previously measured field-strength data and implemented in a field-strength formula based thereon. The accuracy of the method -when compared with about 16,000 measured monthly medians contained in CCIR data bank D- equals that of mainframe computer predictions. The average difference is about 0 dB and the standard deviation is about 11 dB.

A diskette (5 1/4", 360KB, MS-DOS) with the source code and the compiled programs is available from the authors upon request.

1. Introduction

CCIR-Report 340 - the CCIR Atlas of Ionospheric Characteristics - contains the ionospheric characteristics in two alternative forms: the "coefficients" of the numerical mapping functions of the ionospheric parameters for use on computers, and the "charts" which were computed from the coefficients for users not having access to a computer (1).

The Atlas contains 988 numerical coefficients for foF2 and 441 coefficients for M(3000)F2 for each month and for two solar epochs. The storage requirement is (988 + 441) * 12 * 2 = 34296 coefficients. All coefficients are used in any set of calculations. The coefficients are stored as 8 byte per coefficient resulting in a total storage field of 274 KByte.

Alternatively, a grid-point method based on tabular values was suggested which requires spatial and temporal interpolation. This approach, giving a data accuracy commensurate with the CCIR representations, involves storage of values every 5 degrees of latitude and longitude for each two hours. This calls for a storage field of 3.1 MByte.

The large storage capacity needed for the grid method led to the proposal that a procedure based on numerical coefficients be implemented for micro-computer evaluation (Banks et al., 1983).

A completely different approach, however, was applied when the program MINIMUF (3,4,5) was developed, a procedure for calculating the MUF (maximum usable frequency) by means of a simple mathematical model of the ionosphere. This procedure was coded in BASIC for use on personal and home computers, requiring only 3.6 kByte of memory.

2. Accuracy of MINIMUF

Some authors, however, found that the accuracy of MINIMUF is insufficient when compared with measured MUFs and with a mainframe computer MUF prediction (Caruana and Fox, 1985). According to that publication, a mean error of -2.5 MHz and a standard deviation of 6.3 MHz result, when MINIMUF-predicted MUFs are compared with about 15000 measurements of 8 ionosondes. The authors of MINIMUF claim a standard deviation of only 3.8 MHz when MINIMUF-predicted MUFs were compared with 785 measured MUFs obtained on oblique paths of different lengths.

In the course of the work described in the present paper, MINIMUF was compared with the CCIR-Atlas (CCIR, 1986). This represents the internationally agreed basis, recommended by the ITU for the calculation of HF radio propagation predictions.

The prime reason for a sufficiently accurate method of determining the ionospheric parameters is the growing demand for world-wide reliable field-strength predictions. A comparison of MINIMUF with the values obtained from the CCIR-Atlas was made in the following way:

MUF(3000)-values for each point of intersection between the latitudes of 0, +/-10, +/-20, +/-30 etc. degrees (19 numbers) and the longitudes of 0, 30, 60 etc degrees (12 numbers) were calculated for 24 hours of the day, for 12 months and 2 sunspot numbers, yielding a total of 19 * 12 * 24 * 12 * 2 = 131,328 values, which is a much larger sample than that used by (Rose et al., 1978) and by (Caruana and Fox, 1985).

The average difference Σ(MINIMUF - Atlas)/n amounts to 1.0 MHz, indicating that MINIMUF overestimates the MUF(3000) by that amount. The standard deviation

$$RMS = \sqrt{\frac{\Sigma\,(\mathrm{MINIMUF - Atlas})^2 - \frac{(\Sigma\,(\mathrm{MINIMUF - Atlas}))^2}{n}}{n-1}},$$

amounts to about 4.4 MHz, a value similar to the 3.8 MHz stated by the authors of MINIMUF. It was shown (Damboldt and Suessmann, 1988) that most of the deficiencies reported by (Caruana and Fox, 1985) are confirmed, although the magnitude of the error seems to be smaller in the present comparison. It is particularly doubtful whether or not any meaningful systematic latitudinal

variation can be detected with the limited number of observations used by (Caruana and Fox, 1988). Figure 1 shows a histogram of the distribution of differences between MINIMUF- and CCIR-predicted MUF(3000). The large number of cases where the difference is greater than 10 MHz is obvious.

3. A MUF estimation of improved accuracy (FTZMUF2)

The above results demonstrate that MINIMUF, the widely-known MUF prediction for micro-computer use lacks accuracy. This gave rise to the development of a method of improved accuracy, but nevertheless suited for small computers.

The approach chosen is a grid-point method, where the values between the grid points are obtained by linear interpolation. After many trials and comparisons with the CCIR-Atlas, a simple table was obtained which contains the 24 local-time values of foF2 for 0, +/-20, +/-40, and +/-60 degrees geomagnetic latitude. These values were taken from the CCIR-Atlas, calculated for just those grid points. There is no local-time dependence for the latitudes of +90 and -90 degrees, and hence only one number for each Pole is taken into account. The table with the values for March, June, September and December and for two solar epochs (sunspot numbers 0 and 100), comprises 768 numbers only. Actually, each of the 768 numbers in the table is is an artificial merger of both, the value of the northern and the corresponding value southern hemisphere in order to save memory space.

foF2-values for the eight months not contained in the table can be obtained by interpolation. At first, this led to an unacceptably large error, the influence of which was eliminated by means of an additional table correcting the between the values for the interpolated months and the values in the CCIR-Atlas. Similarly, another table with only 96 values (24 hours, 2 latitudes, 2 solar epochs) and a correction table with 20 values were set up for calculating M(3000). All these numbers are compiled in a file "FTZMUF2.DAT" which is the only data source required by the present method in order to calculate the ionospheric characteristics.

The MUF(4000), which is used in propagation predictions, is calculated from foF2 and M(3000) as (CCIR, 1986b):

$$MUF(4000) = 1.1 * foF2 * M(3000)$$

In applying the tables FTZMUF2.DAT, the procedure to obtain a certain MUF-value for a given hour (in UTC), month, location and sunspot number is as follows:

1. calculate local time,
2. change month when on southern hemisphere,
3. determine geomagnetic latitude,
4. look up foF2 for the nearest two latitude belts and the two nearest months,
5. interpolate for latitude and month (including the correction),
6. interpolate for sunspot number, and
7. calculate M(3000) and MUF(4000).

4. The computer program "FTZMUF2" and its accuracy

In order to perform the necessary calculations for estimating foF2 and M(3000) for any location on the earth, a computer program named FTZMUF2 was written (Damboldt and Süssmann, 1988). The language used was Microsoft Basic and the program runs under BASICA as well as under GWBASIC. The program requires 8.7 KByte memory, including code and tables. This is twice the memory required by the MINIMUF code. Computing time for the 24 hourly values of foF2, M(3000) and MUF(4000), however, is approximately only half of that required by MINIMUF.

The accuracy of FTZMUF2 was tested by applying the same method of comparison as described above for MINIMUF. 131,328 values for MUF(3000) were computed using the CCIR-Atlas and FTZMUF2. The average difference Σ(MINIMUF - Atlas)/n amounts to merely -0.09 MHz. The standard deviation is 2.3 MHz which is a substantial improvement upon MINIMUF. Figure 2 depicts the histogram of the distribution of differences between FTZMUF2- and the CCIR-predicted MUF(3000)-values.

It shall be mentioned, that besides calculating the median values (50%-values), the program also allows the calculation of 10%- and 90%-values of foF2 and M(3000) by applying a simplification of the distribution given in (CCIR, 1986b).

5. Input parameters for FTZMUF2

The computer program asks for the following input parameters to be specified:

* geographical latitude and longitude (North and East positive, South and West negative) e.g. the coordinates for Darmstadt are entered as: 49.8,8.6
* month (Jan = 1, Feb = 2 etc),
* sunspot number, and
* percentage of time wanted (10, 50 or 90).

6. Output parameters of FTZMUF2

The program outputs a table of the 24 hourly values of

* foF2,
* M(3000) and
* MUF(4000)

for the specified percentage of time.

7. FTZ HF prediction model

The FTZ field-strength prediction method involves four basic steps:

- determination of the basic MUF (see program FTZMUF2)
- determination of the upper frequency limit f_M
- determination of the lower frequency limit f_L and
- an estimate of the variation of field strength with frequency within these limits.

The details of these steps are discussed in the following sections.

8. Determination of the basic MUF and the elevation angle

A control-point method is applied to determine the basic MUF for oblique paths of length D. For single-hop pat: · (for E-layer less than 2000 km, for F-layer less than 4000 km) the control point is at the path midpoint. For multi-hop paths the . ntrol points are the midpoints of the first and last hops.

For long circuits, particularly for paths through the auroral zone, it has been found necessary to take into account a third control point at the path midpoint. Therefore, three control points are considered for multi-hop circuits.

The hop length is calculated by dividing the great circle distance into hops of a maximum length of 2000 km for the E-layer and into hops of a maximum length of 4000 km for the F2-layer propagation. The elevation angle is determined for the given hop length and a reflection height of 110 km (E-layer) and 300 km (F2-layer), respectively. If the elevation angles are below the wanted minimum elevation angle the number of hops is increased by one and the calculation of elevation angles is repeated until the wanted minimum elevation angle is exceeded.

The 24 hourly values of the F2-layer (F2(4000)MUF) are determined in the subprogram FTZMUF2 described in the first part of this paper. These 24 hourly values are converted into the F2 basic MUF by:

$$\mathrm{F2(D)MUF} = \mathrm{F2(4000)MUF} \cdot F_D$$

with $F_D = ((((((-6.713 \cdot 10^{-9} \cdot \mathrm{Farc} + 4.492 \cdot 10^{-7}) \cdot \mathrm{Farc} - 9.986 \cdot 10^{-6}) \cdot \mathrm{Farc} + 6.865 \cdot 10^{-5}) \cdot \mathrm{Farc} + 9.202 \cdot 10^{-5}) \cdot \mathrm{Farc} + 0.002265) \cdot \mathrm{Farc} + 0.004699) \cdot \mathrm{Farc}$

where
Farc = F2-hoplength in radians

The critical frequencies f_c (E) for the E-layer are determined as follows (Rawer 1953):

$$f_c(E) = K_E \cdot \cos^n \chi$$

$$K_E = 2.25 + 1.5 \cos\varphi + (0.01 - 0.007 \cos\varphi)\, R_{12} \quad \mathrm{(MHz)}$$

$$n = 0.21 + 0.12 \cos\varphi + 0.0002\, R_{12}$$

where
R_{12} = 12-month running mean of the sunspot number

χ = sun's zenith angle

φ = geographical latitude.

The 24 hourly values of the E-layer (E(D)MUF) are determined by:

$$\mathrm{E(D)MUF} = f_c(E) \cdot 5 \cdot E_D$$

with $E_D = ((((((-4.368 \cdot 10^{-8} \cdot \mathrm{Earc} + 1.335 \cdot 10^{-7}) \cdot \mathrm{Earc} - 5.977 \cdot 10^{-6}) \cdot \mathrm{Earc} + 0.0002625) \cdot \mathrm{Earc} - 0.005039) \cdot \mathrm{Earc} + 0.03761) \cdot \mathrm{Earc} - 0.01332) \cdot \mathrm{Earc} + 0.2085$

where
Earc = E-hoplength in radians.

The basic MUF for the whole circuit is chosen by taking the higher value of the E- or F2-MUF for each control point and then taking the lowest value of all three control points.

9. Determination of the upper frequency limit, f_M

The basic MUF denotes the highest frequency that can be propagated by a particular mode between specified terminals by ionospheric reflection. However, experience has shown that reception is possible above the basic MUF. The receiving field strength does not abruptly fall off when the basic MUF is exceeded (Damboldt and Suessmann, 1976). The field strength decreases gradually with increasing frequency, therefore the "operational MUF" increases with increasing transmitter power. In most cases the operational MUF can be substantially higher than the basic MUF. This is the consequence of several mechanisms which are not taken into account by prediction techniques based on theoretical considerations of effective propagation modes. The efficiency of these mechanisms increases with increasing beamwidth of the transmitting and receiving antennas.

Since - at present - it is not possible to quantitatively determine all influences by adequate formulae an empirical correction factor K is applied to each of the aforementioned 24 hourly values of the basic F2(D)MUF in order to yield the "upper frequency limit" f_M (which here is a special case of the "operational MUF" as defined in (CCIR, 1976a).

$$f_M = K \cdot f_g \cdot C_r$$

$$\text{with } K = 1.2 + W \frac{f_g}{f_{g,noon}} + X \left(\left(\frac{f_{g,noon}}{f_g} \right)^{1/3} - 1 \right) + Y \left(\frac{f_{g,min}}{f_{g,noon}} \right)^2$$

where

f_M : upper frequency limit for a transmitter power of 1 MW e.r.p. and a receiving field strength of 1 µV/m (or 1 kW e.r.p. and 30 dB below 1 µV/m),

f_g : hourly value of F2(D)MUF,

D : ground range between terminals, km,

$f_{g,noon}$: f_g for local noon of the respective control point,

$f_{g,min}$: minimum value of the 24 hourly values of F2(D)MUF,

C_r : $2 - \left(\frac{D}{4000}\right)^2$ (for circuits < = 4000 km) or 1 (for circuits > 4000 km)

W, X, Y : constants

The constants W, X, Y are chosen empirically, they depend on the geographical position of the radio circuit. To a certain degree they take into account the influences on propagation by the auroral zone and the geomagnetic equator. For east-west and north-south paths the numerical values of these constants are shown in Table I.

Table I: Empirical values for W, X, Y

	W	X	Y
east-west	0.1	1.2	0.6
north-south	0.2	0.2	0.4

Values for circuits which are not exactly east-west or north-south are obtained by linear interpolation of the numbers in Table I depending on the bearing at the path midpoint.

The variability of the first two terms in the formula of the K-factor represents the ratio of the hourly value of the basic MUF f_g to the noon value $f_{g,noon}$ and to its reciprocal value, respectively. The term with the coefficient X describes the increase of the K-factor from day to night. However, since the K-factor has to reach a certain value during daytime, the term with the coefficient W is added. It counteracts the term mentioned first. The last term with the coefficient Y varies with the ratio of the minimum basic MUF $f_{g,min}$ to the noon value $f_{g,noon}$ and expresses the variability of the K-factor from winter to summer.

The physical ideas behind the formula for the K-factor are as follows: The dynamics of the ionosphere manifest itself in the diurnal and seasonal variations of the basic MUF. On the other hand ionospheric irregularities become the more apparent the more the height of the F-layer increases and its critical frequency decreases. Consequently, the highest value for K is obtained in winter nights when the basic MUF values are at their lowest.

10. Determination of the lower frequency limit, f_L

After the upper limit of the transmission frequency range has been determined the ionospheric absorption is computed which influences the lower limit of the transmission frequency range. Towards the lower frequency end of the HF-range the non-deviative D-layer absorption plays a decisive role. Therefore, the lower frequency limit is calculated with the aid of the formula for non-deviative absorption:

$$L = \frac{B_o \sum_{2N} \sqrt{\cos \chi} \cdot (1 + 0.005 \cdot R_{12})}{(f + f_H)^2 \cos \varphi_D}$$

where

χ : sun's zenith angle,

B_o : constant which comprises the basic values for the average collision number and ionization density and the average thickness of the D-layer

φ_D : angle of incidence at the D-layer

f_H : the gyrofrequency

N : Number of hops

The lower frequency limit f_L ,which is defined here as the frequency where the field strength of a 1 MW transmitter is 1 µV/m (same definition as for the upper frequency limit f_M) ,is now calculated as follows:

$$f_L = 5.3 \left[\frac{\sum_{2N} \cos^{1/2} \chi \ (1 + 0.009 \ R_{12})}{\cos \varphi_D \ \ln \frac{9.5 \cdot 10^6}{D_p}} \right]^{1/2} - f_H$$

where

D_p : oblique path length.

With the above formula for f_L, the lower frequency limit is determined for the daylight hours. During the night the lower frequency limit is assumed to be dependent only on the distance between transmitter and receiver:

$$f_{LN} = \left(\frac{D}{3000} \right)^{1/2}, \text{MHz}$$

For a distance of D = 3000 km the night-LUF becomes 1 MHz. As there is a certain lag between the time of sunset and the decrease of region ionization, the decay from day-LUF to night-LUF is assumed:

- to last for three hours,
- to begin at double the night-LUF and
- to follow an exponential law.

During this three-hour interval the LUF is obtained as follows:

$$f_L = 2 \ f_{LN} \ e^{-0.23t}$$

After the upper and lower limits of the transmission frequency range have been determined, the next step is to compute the field strength as a function of the frequency within the limits given by f_M and f_L.

11. Variation of field strength with frequency

Field strength recordings made by the Deutsche Bundespost for many years clearly showed a systematic variation of field strength within the transmission frequency range such that the field strength steadily increases from the lower frequency limit to higher frequencies, following approximately an inverse frequency dependence. After a maximum value is reached the field strength decreases steadily until it reaches the upper frequency limit. Already on the descending part of the curve - after the maximum - we can see the basic MUF, up to which propagation by ionospheric refraction takes place. Above the basic MUF propagation by off-great-circle- and scatter mechanisms occur.

As a consequence of these observed facts an empirical formula was developed by Beckmann (1965). This formula is based on the following expression for the field strength E which is valid for the non-deviative absorption:

$$E = E_0 \left(1 - \frac{f_L^2}{f^2} \right),$$

where E_0 is the free-space field strength given by:

$$E_0 = 20 \log \frac{173\,000 \sqrt{P}}{D} \quad \text{dB} (\mu\text{V/m})$$

with P as the effective radiated power (e.r.p.) in kW and D as the distance between transmitter and receiver in km. In the case of P = 1 MW, E_0 becomes

$$E_0 = (139.6 - 20 \log D) \quad \text{dB} (\mu\text{V/m})$$

As the deviative absorption increases with increasing frequency, the term

$$\frac{f^2}{f_M^2}$$

has to be added to the above formula for E. Then a factor C has to be taken into account to meet the conditions that E vanishes for $f = f_L$ and $f = f_M$:

$$C = \frac{1}{1 + \left(\frac{f_L}{f_M}\right)^2}$$

The Beckmann formula for the variation of field strength within the transmission frequency range is then obtained as:

$$E = E_0 \left(1 - \frac{f_M^2}{f_M^2 + f_L^2} \left(\frac{f_L^2}{f^2} + \frac{f^2}{f_M^2} \right) \right) - 30.0\ \text{dB} + G_t + FG + 10 \log (P) \quad \text{dB} (\mu V/m)$$

with

$$FG = 10 \log \frac{C}{\sin (C)} \qquad FG\ \max. = 15\ \text{dB}$$

$$D <= 20\,000\ \text{km} : C = A$$

$$D > 20\,000\ \text{km} : C = 2 \cdot \pi - A$$

$$D > 30\,000\ \text{km} : C = A - \pi$$

$$A = \frac{D \cdot \pi}{20\,000\ \text{km}}$$

where

P = Transmitter-power in kW

G_t = Transmitter-antenna gain in dBi

where, again, E_0 is the free-space field strength (in dB above 1 μV/m). In the above formula for the field strength, 30 dB are subtracted because in practice the transmitter powers are normalized to 1 kW e.r.p. and the field strength therefore is 30 dB below the value computed for 1000 kW.

For a more exact computation of the field strength the influence of the Earth's magnetic field has to be allowed for. In practice it has proved to be sufficient to add the value of the gyrofrequency (about 1 MHz) to the frequencies f, f_L and f_M.

It should be stressed that the Beckmann formula empirically comprises all different factors influencing propagation. Modes, polarization coupling loss, blanketing and other phenomena are not dealt with separately as is the case in the analytical field-strength prediction methods. These phenomena are partly contained in the values characterizing the transmission frequency range, namely in f_L and f_M. The empirically determined K-factor of the MUF computation comprises all other influences on the field strength not yet determined quantitatively.

It has been pointed out (Damboldt and Süssmann, 1985) that at short distances the operational MUF can be substantially higher than f_M leading to higher field strength than those calculated as above. This is taken into account by an additional correction factor, increasing the field strength at frequencies above the basic MUF and at shorter ranges (up to 4000 km). This correction factor is a function of distance and of the ratio f/basic MUF. It is zero for distances > = 4000 km, thus leading to the "pure" long-distance method as described above.

12. Input parameters for MINIFTZ4.2

As input, the program requires:
* month
* year
* sunspot number
* Tx-name
* Tx-coordinates
* RX-name
* RX-coordinates
* TX-power
* TX-antenna gain (dBi)
* percentage of time
* short/long path
* option for angle/mode output
* minimum angle
* frequencies for which field strength is predicted (max. 11)

13. Output parameters of MINIFTZ4.2

The program calculates hourly values of the basic MUF, the field strength at the basic MUF, the FOT (frequency of optimum traffic, here taken as 90 percent of the basic MUF) and the field strength for up to 11 frequencies. The program has the option to output modes and elevation angles producing a similar table as that for field strength. For each frequency and hour the number of hops, whether E- or F- reflection occurs and the elevation angle (in degrees) are indicated. No modes and elevation angles are given when the corresponding field strength is below -40 dB (1 μV/m) or when the frequency is higher than the 10-percent basic MUF.

14. References

BANKS, P.H.T., B.H. MILLER, M.I. DICK and P.A. BRADLEY (1983) Storing of the CCIR reference maps of ionospheric characteristics for microcomputer evaluation, CCIR IWP 6/1 Doc 244 (29 July 1983)

BECKMANN, B. (1965) Bemerkungen zur Abhängigkeit der Empfangsfeldstärke von den Grenzen des Übertragungsfrequenzbereiches (Remarks on the receiving field strength as a function of the limits of the transmission frequency range), NTZ, Vol. 19, 643-655.

CARUANA, J. and M.W. FOX (1985) A critical study of the NOSC HF Prediction Alorithm MINIMUF, Technical Report IPS-TR85-07, Department of Science, IPS Radio and Space Services Australia, June 1985

CCIR, Dec. 373-5 (1986a) Definitions of maximum transmission frequencies CCIR Geneva

CCIR, Report 340 (1986b) CCIR-Atlas of ionospheric characteristics, CCIR Geneva

DAMBOLDT, T. and SÜSSMANN, P. (1976) Extending field strength predictio ns for short distances to frequencies above the standard MUF. IWP 6/1 Doc. 45. Translated from Kleinheubacher Berichte Vol. 19, 527-432

DAMBOLDT, T. and SÜSSMANN, P. (1985) Proposal for MUF-dependent Ly and Lz-values, CCIR IWP 6/13 Doc 7 (1 April 1985)

DAMBOLDT, T. and SÜSSMANN, P. (1988) A simple method of estimating iono spheric parameters foF2 and M(3000) with the aid of a home computer. Technischer Bericht 445 TB 25 E, December 1988, FTZ D43-DrV, P.O. Box 5000, D-6100 Darmstadt

RAWER, Karl (1953) Die Ionosphäre, P. Noordhoff N.V., Groningen Holland , p. 176

ROSE, R.B., J.N. MARTIN and P.H. LEVINE (1978) MINIMUF-3: A simplified Hf MUF Prediction Algorithm, Naval Ocean Systems Center, San Diego, USA, Technical Report TR 186, 1. Feb. 1978

ROSE, R.B. (1982a) MINIMUF-3.5: Improved version of MINIMUF-3, a simplifed hf MUF prediction algorithm, Naval Ocean Systems Center, San Diego, Technical Document TD 201, 26. Oct. 1978

ROSE, R.B. (1982b) MINIMUF: A simplified MUF-Prediction Program for Microcomputers, QST, Dec. 1982, p 36-38

Annex 1

HF-FIELD STRENGTH ESTIMATED BY MINIFTZ4.2

CIRCUIT :	TEHERAN - NORDDEICH	MONTH :	APR. 86
LOCATION:	35.5N 51.3E 53.6N 7.1E	SSN :	7.
AZIMUTH :	314.6 DEG 102.4 DEG	POWER :	10.0 KW
DISTANCE:	3952 KM	ANT GAIN:	12.0 DB
MIN-ANG.:	3.0 DEG.		

FIELD STRENGTH IN DB ABOVE 1 UV/M FOR 50 PERCENT OF TIME

UTC	MUF	DBU	FOT	3.0	4.0	5.0	6.0	8.0	10.0	12.0	15.0	18.0	22.0	26.0
1	7.1	27	5.2	32	35	35	32	22	7	-11	...	...	...	...
2	6.9	27	5.1	31	35	34	31	20	5	-14	...	...	...	...
3	6.7	25	5.0	25	30	31	28	18	2	-17	...	...	...	...
4	7.1	17	5.4	-13	6	15	18	15	4	-11	-40	...	...	...
5	8.2	13	7.9	...	-16	1	9	14	9	-1	-22	...	...	...
6	9.8	14	9.7	...	-31	-9	4	14	14	8	-7	-27	...	...
7	11.1	15	10.8	...	...	-16	-1	13	16	13	3	-13	...	...
8	12.0	15	11.6	...	...	-21	-4	12	17	15	7	-6	-30	...
9	12.6	16	12.1	...	...	-24	-6	11	17	17	9	-3	-25	...
10	13.0	16	12.4	...	...	-25	-7	11	18	18	11	0	-21	...
11	13.4	17	12.4	...	...	-24	-6	12	19	19	13	2	-18	...
12	13.6	17	12.3	...	...	-20	-3	14	20	20	14	3	-17	...
13	13.6	18	11.6	...	...	-14	1	17	22	21	15	3	-17	...
14	13.5	19	10.0	...	-29	-5	8	21	24	22	15	3	-18	...
15	13.1	21	9.7	...	-9	8	18	26	27	24	15	2	-21	...
16	12.8	24	9.3	-2	17	26	31	34	32	26	15	0	-26	...
17	12.1	25	8.6	16	28	34	37	36	32	25	12	-6	-34	...
18	11.2	26	7.7	25	34	38	39	37	31	23	7	-13	...	...
19	10.1	27	7.1	31	37	40	39	35	28	18	-1	-25	...	...
20	9.2	27	6.6	34	38	39	38	32	23	11	-12	-39	...	...
21	8.6	27	6.2	33	38	38	37	30	19	5	-20	...	...	...
22	8.0	27	5.8	33	37	37	35	27	15	0	-28	...	...	...
23	7.7	27	5.5	32	36	36	34	25	13	-4	-33	...	...	...
24	7.4	27	5.3	32	36	35	33	23	10	-8	-39	...	...	...

MODES AND ELEVATION ANGLES ESTIMATED BY MINIFTZ4.2

CIRCUIT:	TEHERAN - NORDDEICH	MONTH :	APR. 86
LOCATION:	35.5N 51.3E 53.6N 7.1E	SSN :	7.
AZIMUTH :	314.6 DEG 102.4 DEG	POWER :	10.0 KW
DISTANCE:	3952 KM	ANT GAIN:	12.0 DB
MIN-ANG.:	3.0 DEG.		

MODES AND ELEVATION ANGLES

UTC	MUF	MODE	FOT	3.0	4.0	5.0	6.0	8.0	10.0	12.0	15.0	18.0	22.0	26.0
1	7.1	2F13	5.2	2F13	2F13	2F13	2F13	2F13	2F13	....	....	....	....	....
2	6.9	2F12	5.1	2F12	2F12	2F12	2F12	2F12	....	....	....	....	....	....
3	6.7	2F12	5.0	2F12	2F12	2F12	2F12	2F12	....	....	....	....	....	....
4	7.1	2F12	5.4	2F12	2F12	2F12	2F12	2F12	....	....	....	....	....	....
5	8.2	2F12	7.9	....	3E06	3E06	3E06	2F12	....	....	....	....	....	....
6	9.8	2F12	9.6	....	3E06	3E06	3E06	3E06	2F12	....	....	....	....	....
7	11.1	2F12	10.7	....	....	3E06	3E06	3E06	3E06	2F12	....	....	....	....
8	12.0	2F12	11.4	....	....	3E06	3E06	3E06	3E06	2F12	....	....	....	....
9	12.6	2F12	11.8	....	....	3E06	3E06	3E06	3E06	2F12	....	....	....	....
10	13.0	2F12	12.1	....	....	3E06	3E06	3E06	3E06	3E06	....	....	....	....
11	13.4	2F12	12.1	....	....	3E06	3E06	3E06	3E06	3E06	2F12	....	....	....
12	13.6	2F12	12.0	....	....	3E06	3E06	3E06	3E06	3E06	2F12	....	....	....
13	13.6	2F12	11.5	....	....	3E06	3E06	3E06	3E06	2F12	2F12	....	....	....
14	13.5	2F11	10.0	....	3E06	3E06	3E06	3E06	2F11	2F11	2F11	2F11	....	....
15	13.1	2F11	9.7	....	3E06	3E06	3E06	2F11	2F11	2F11	2F11	2F11	....	....
16	12.8	2F11	9.3	2F11	2F11	2F11	2F11	2F11	2F11	2F11	2F11	2F11	....	....
17	12.1	2F12	8.6	2F12	2F12	2F12	2F12	2F12	2F12	2F12	2F12	....	....	....
18	11.2	2F12	7.7	2F12	2F12	2F12	2F12	2F12	2F12	2F12	2F12	....	....	....
19	10.1	2F12	7.1	2F12	2F12	2F12	2F12	2F12	2F12	2F12	2F12	....	....	....
20	9.2	2F12	6.6	2F12	2F12	2F12	2F12	2F12	2F12	2F12	....	....	....	....
21	8.6	2F12	6.2	2F12	2F12	2F12	2F12	2F12	2F12	2F12	....	....	....	....
22	8.0	2F12	5.8	2F12	2F12	2F12	2F12	2F12	2F12	....	....	....	....	....
23	7.7	2F13	5.5	2F13	2F13	2F13	2F13	2F13	2F13	....	....	....	....	....
24	7.4	2F13	5.3	2F13	2F13	2F13	2F13	2F13	2F13	....	....	....	....	....

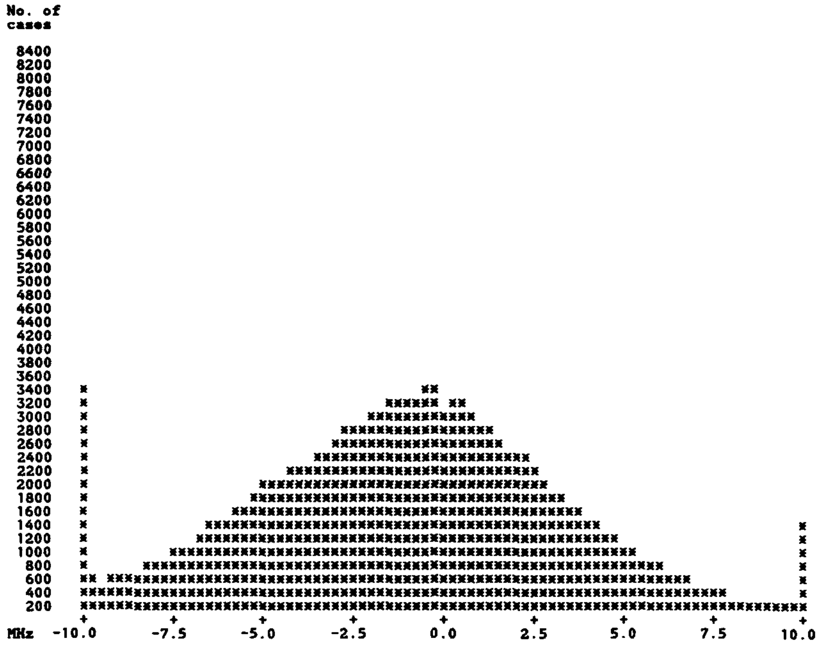


Figure 1: Histogram of the distribution of differences between MINIMUF and CCIR-Atlas for MUF(3000) count: 131328 mean difference: 1.0 MHz standard deviation: 4.4 MHz

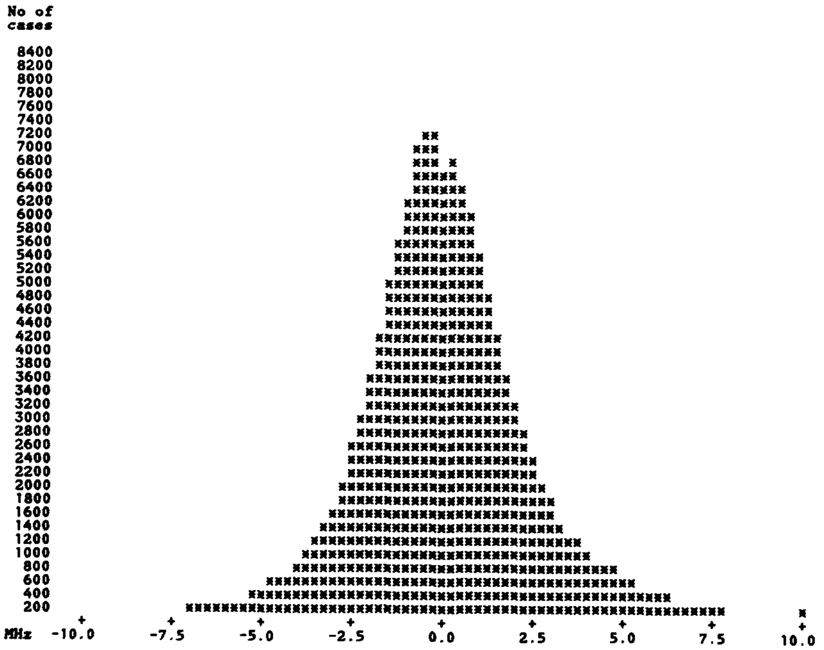


Figure 2: Histogram of the distribution of differences between FTZMUF2 and CCIR-Atlas for MUF(3000) count: 131328 mean difference: -0.09 MHz standard deviation: 2.3 MHz

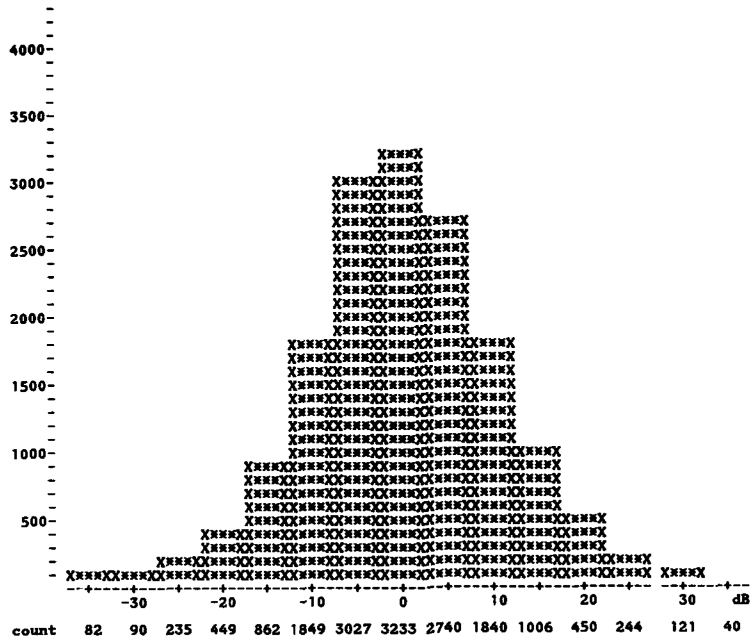


Figure 3: Histogram of MINIFTZ4 predicted field strength versus CCIR Data Bank D1
count: 16268 mean difference: 0.0 dB standard deviation: 10.9 dB

DISCUSSION

C. GOUTELARD, FR
English Translation

The measurements you have presented show a standard deviation of 10 dB between modelled and measured data. I assume that they include disturbed days. When it comes to utilization, it is possible to correct the monthly values by using short term forecasts. A system of this kind has been developped by the CNET in a study conducted by Ms. Juy. Is it possible, for operational utilization, to include such data in your model?

AUTHOR'S REPLY

The monthly medians (whether measured or predicted) are not influenced very much by a small number of disturbed days. I don't therefore believe that the removal of disturbed days would actually have any influence on the standard deviation of about 10 dB. Nevertheless, the main step to be taken in HF prediction methods is certainly the reliable forecast of disturbances.

THE CHARACTERISTICS OF THE HF RADIO CHANNEL AND ITS INFLUENCE ON MODERN HF DATA COMMUNICATION SYSTEM DESIGN

by

Detlev Borgmann and Dr Jürgen Lindner
AEG Aktiengesellschaft
Radio and Radar Systems Group
Postfach 1730
Sedanstrasse 10
D-7900 Ulm/Donau
West Germany

SUMMARY

A basic requirement for the design and realization of modern HF data communication systems is the detailed analysis and modelling of the transmission medium. The paper describes, how this analysis and modelling of the HF radio channel influences the design of such data communication systems.

The first part of the paper is concerned with a description of the characteristics of the HF radio channel by parameters like multipath and Doppler spread and its variation with time, short and long term statistics, variation of useful frequencies with time and it gives a short introduction into modelling.

The second part of the paper then shows, in which way a modern system design can cope with the unwanted characteristics of the transmission medium. As an example a system is presented, which is realized according to the open systems interconnection architecture proposed by ISO and in which the protocols of layers 1 to 3 are adapted to the medium HF radio.

The third part of the paper describes details of an adaptive frequency management system with its operational requirements, the basic functions and the position within the communications system. The frequency management system is implemented as a functional modul of a radio link processor, which is a part of the data communication system. The frequency management system comprises the following features:

a) long term prediction and analysis
b) short term analysis with link statistics
c) channel monitoring with measurement of noise and interference

Some results of field test with this frequency management system are presentet.

1. INTRODUCTION

Modern HF communication system design offers the chance to cope with the known disturbing effects of the ionospheric transmission medium. New techniques and equipment open this traditional long range communication medium to modern data communication applications.

Some work was published in the past (see, e. g. /1/) but only a few papers deal with entire HF data communication systems /2/. On the other hand it is quite evident, that only the communication system approach, e. g. on the basis of the ISO/OSI modell for the interconnection of open systems, can give satisfactory results.

The background of this paper is the design an the realization of a data communications system, which the German army intends to introduce in the near future.

A basic requirement for the design of fast and reliable HF data communication systems is a profound knowledge of the HF transmission medium and its disturbing effects. The following section is devoted to this topic.

2. *THE HF RADIO CHANNEL*

2.1 Disturbing effects

Figure 2-1 shows the first of three disturbing characteristics which affect HF radio transmissions: multipath propagation. This means that a transmitted signal can reach the receiver via several paths at the same time. This effect is particularly annoying if high-speed data are transmitted (e. g. 3 kbit/s), since a transmitted data symbol reaches the receiver, e. g. a second time, a few data symbols later, thus interfering the data symbol being received at that time. It was due to this effect of "intersymbol interference" that conventional data transmission via HF radio uses signalling rates less than 200 baud (typical 50 or 75 baud).

The second disturbing effect is that the frequencies which can be used for transmission vary with the position of the sun and solar activity. Figure 4-1 shows the typical profile of useful frequencies over a period of 24 hours. The most striking feature is the "narrowing" of the band at night. Only a relatively narrow frequency band can be used during night time. This characteristic of the ionosphere has always meant - and still does - that the skill and know-how of the radio operator, as well as careful frequency planning, is of decisive importance for conventional transmissions via HF radio.

The profile of the useful frequency band limits shown in this picture was determined with the aid of the ionospheric propagation prediction program IONCAP* /3/. The input parameters for this program were the coordinates of the transmitter and receiver (Flensburg and Ulm), the transmitter power (500 W), the antenna types (horizontal dipoles), the month (December) and the number of sun spots (200), as well as a required signal-to-noise ratio per Hz bandwidth (45 dB/Hz). The profile predicted by IONCAP correlates relatively well with the situation in reality. Thus, such methods are suitable for automatic (on-line) selection of radio frequencies. This will be described in more details in section 4.

A third group of disturbing effects consists of additive interference. These include not only noise, interference, atmospherics and man made noise, but mainly the interference caused by other HF radio users.

2.2 Modelling

The disturbing effects observed in reality are of time-varying stochastic nature. A statistical description leads to non-stationary stochastic processes, i. e. the statistical description is a function of time.

For a given radio frequency and a given bandwidth (e. g. 3 kHz) the multipath propagation and the additive interference have to be modelled. Usually this is done in the following way:

a) *modelling of short term variations:*
   stationary statistics (time invariant)

b) modelling of long term variations:
   cyclostationary statistics (periodically time dependent)

Short term variations mean variations in the range of seconds or some 100 milliseconds. The Watterson model /4/ is a very useful approach to model the short term variations of multipath propagation.

The upper part of fig. 2-2 is a description of the observations and the lower part shows the corresponding modelling according to the Watterson model. The multiplicative complex pass band noise consists of two independent Gaussian noise processes (i. e. Rayleigh statistics) with the following spectral parameters (each process):

- center frequency	=	Doppler shift
- bandwidth	=	Doppler bandwidth
- total signal power	=	square of rms-value

To model the entire multipath propagation some of those single path models have to be combined, with their individual time delay and spectral parameters. Fig. 2-3 describes the resulting model for multipath propagation. Each of the complex gain factors is controlled by two independent Gaussian noise processes which are also independent of the others.

One means to illustrate the resulting time-varying behaviour of multipath propagation is the so-called fadeogram - see fig. 2-4. At the horizontal axis is the time in seconds and the vertical axis represents the frequency in kHz relative to the HF center frequency. The grey-level represents the transfer function. White means good transmission, black means a fade out. The first two fadeograms (upper part) of fig. 2-4come from a Watterson channel simulator (with two sets of parameters) and the other two fadeograms are taken from measurements on a 800 km real HF link. Time-varying frequency selective fading - that ist the counterpart of time-varying multipath propagation in the frequency domain - is obvious.

The parameters of the time-varying multipath propagation (i. e. number of paths, the individual delays, Doppler shifts, Doppler bandwidths, rms values) are quantities which belong to the long term varying category.

For the additive interference a variety of signals are used for modelling: CW, noise, ASK, FSK, PSK, analog speech signals. The special type and the power of the interfering signals can also thought to be long term varying.

Fig. 2-5 shows the short term model of the HF radio channel. It ist a very useful tool for the development of HF modems.

* INOCAP = Ionospheric Communications Analysis and Prediction Program

The long term variations (of the short term parameters) are commonly modelled in a very rough way. Neither the detailed parameters of multipath or the type of interference are considered, only signal and interference (noise) power. This may be justified as follows: if the data transmission method (i. e. the HF modem) has been designed properly, only signal and noise power have an influence on data bit error rate. The dependence on the remaining parameters of multipath and interference can be neglected.

For given locations of transmitter and receiver, transmission power and antenna diagrams simulation programs like IONCAP /3/ may be taken to predict average signal and average interference power at a given time of day, month, year.*
The dependence of these averages on the HF radio frequency is also predicted. With fig. 4-1 this dependence has been described before.

3. HF RADIO SYSTEM DESIGN

3.1 Architecture

For the design of an HF data communication system the following aspects have to be considered:

a) the HF transmission medium and its influence on the service required

b) the required service and the operational requirements

A modern design approach has to leave only those activities to the user, which are really neccesary. To illustrate this, an example will be considered in the following. This example is the transmission of telex messages trough an HF radio network.

The approach is, to leave only the preparing of the telex message at the user side and to put all other activities into the system, i. e. fast and error free transmission of the telex message to their destination address.

To have a transparent design, it is useful to take a model, with which the necessary functions and protocols to realize can be structered. Fig. 3-1 shows the seven layer model of the ISO for the open system interconnection (OSI). This model will be used in the following.

Fig. 3-2 shows, how the layers (or functions) can be connected with real equipment. Layer 1 comprises the HF modem and the HF radio equipment, layers 2 and 3 are contained in the "radio link processor".

The functions and protocols of layers 1 to 3 depend on the HF transmission medium (a) above), whereas those of layers 4 to 7 do not depend on the HF transmission medium, but on the required service and the operation by the user. Layers 4 to 7 are the application specific layers.

In the following the specific functions of layers 1 to 3 are considered.

3.2 The HF modem

It is known that data transmission can be made as effective as possible only, if a transmission method is used, which is adapt[illegible] the channel an its actual properties. This means for the HF radio channel and the development of modems that we need the short term model explained in the last section.

Because of the time-varying nature of the HF radio channel it is necessary to use adaptive methods at the receiving side. The adaptation methods must enable the modem receiver to track every variation of the channel.

Some requirements must be met with a state of the art HF modem:

- The bandwidth should be as small as possible
- The envelope of the transmitted waveform should vary as little as can be
- The average bit error rate should be as small as possible also in case of multipath propagation and strong interference

* averages are ensemble averages; practically they can be taken over different periods of the cyclo-stationary time varying statistics; the period may be e. g. one 11 years sun spot number cycle. Usually the period is taken to be one day and the remaining long term parameters (months, year, sun spot number, interference level) are input parameters determined by the user of the prediction program.

The first requirement leads to linear modulation methods like PSK, ASK (generally QAM) and the second leads to single-tone or serial transmission formats.

This is in contrast to conventional parallel-tone formats where 16 or more tones are modulated and transmitted at a slow rate in a frequency division like manner. Here a loss in HF power of 8 dB or more must be taken into account.

The third requirement leads to coherent reception and to adaptive equalization at the receiving side. It ist known that the decision feedback equalizer is well suited for multipath channels like HF radio. Decision feedback equalizers have been used for data transmission via analog telephone circuits. In this case the methods for the equalizer adaption can be simple. This is in contrast to HF, where fast and robust adaptation algorithms are necessary. Beside this a structure of data transmission is required, which allows fast recovery after flat fadings (i. e. the entire input signal vanishes) or strong interference.

Fig. 3-3 shows an example of features of an HF modem, which result, if all these techniques are applied.

Fig. 3-4 shows the bit error rate which results, if data are transmitted with this modem over a channel with time-varying multipath and additve noise interference. To get reproduceble results, this transmission was made with the HF channel simulator explained before. For comparision the curve for a conventional parallel tone type modem (39 tones) is shown too. The obvious difference can be seen at low bit error rates: to get a bit error rate of about 10^{-3} e. g. , the parallel tone modem needs about 15 dB more HF power.

3.3 The radio link processor

On HF radio it is nearly impossible to have acceptable conditions on one frequency during all time. The frequency must be changed from time to time.

There are some reasons:

- the variation of the ionospheric layers in the course of a day (see fig. 4-1, long term variation)
- strong co-channel interference
- jamming

To cope with difficulties like these, usually a radio operator observes the transmission and changes frequencies if necessary. Beside this, the operator is also responsible for link set-up.

Functions like these can be handled in a far more effective way by fast processors with properly designed protocols. A necessary pre-condition is, that more than one frequency is available. In the case of HF radio networks another aspect becomes important: the most effective use of a common pool of frequencies assigned to the network, in order to maximize the throughput within the network. The pool of frequencies has to be managed like a bundle of transmission lines in a telephone network.

In this way the radio link processor must have an internal "adaptive" frequency management, which takes into account a wide variety of parameters, e. g. actual error rate, HF propagation prediction, knowledge gained form previous transmissions,occupation by other network users, jamming. The adaptive frequency management is the basis for a proper automatic selection of frequencies in any situation on the HF transmission medium and within the radio network. It will be described in more detail in section 4.

Beside the adaptive frequency management the protocol of the radio link processor plays an important role. Its task is to manage the transmission of data packets under all conditions within the network and with respect to the transmission medium. In every case it must act and react in such way, that the transmission task is performed as good as possible. To get such a protocol a proper design is necessary and an extensive simulation of this protocol within the network - before being implemented in the radio link processor.

This protocol uses the service of the physical layer (layer 1, bit transmission HF modem). To do this effectively the task of layer 2 must be observed, i. e. data protection. In case of HF radio channels this layer 2 comprises techniques like FEC (forward error correction) and ARQ (automatic repeat request).

For HF radio it is very important to have a code which is able to correct random errors and burst errors equally well. The Reed Solomon (RS) code /5/ is a very good approach - in theory and HF radio praxis. The RS code is a block code, based

on 8 bit symbols (not on bits like binary codes) and it can correct t erroneous 8 bit symbols (without regard to the distribution within the code word), where t is half the number of redundancy symbols. This is in contract to binary codes, where t is less than a quarter of the number of redundancy bits.

Fig. 3-5 shows how RS coding improves the bit error rate. For a channel with only additive white Gaussian noise (WGN) there is a sharp cut-off at about 9 dB. If the SNR is increased a little bit, practically no error can be measured any longer. A similar behaviour can be seen for channels with frequency selective and flat fading.

The HF radio channel has a time-varying channel capacity, i. e. the maximum data rate which is possible depends on time. To take care of this effect, the actual transmission rate must adaptate to the actual conditions on the channel. One means to achieve this, is using ARQ. Most effective methods employ a combination of ARQ and FEC.

If layer 2 with FEC and ARQ is contained in the radio link processor, then it is straightforward to use code word lengths and redundancy in a different way: seperate code words with great redundancy for control information and code words and redundancy which are adapted to the data transmission task. For ARQ, e. g. it is sufficient to take about 25% redundancy and for short control words 75% redundancy is a good approach.

75% redundancy means that 37,5% of the code symbols may be in error without effect on the transmission of the control information. In a 3 kHz bandwidth this is equivalent to about 2.5 dB in SNR.

4. ADAPTIVE FREQUENCY MANAGEMENT

4.1 Principles

The main prerequisite to establish a satisfactory HF communication link is the choice of a useful frequency at a given time. The task of selecting the best frequency according to the prevailing conditions is usually described by the term "frequency management". The main criteria which are used today by such a frequency management are

a) acceptable reliability with respect to ionospheric propagations conditions
b) required signal to noise or interference ratio (SNR)
c) tolerable dispersion (multipath effects)

A succesful frequency management depends upon the ability to predict and measure the above listed criteria.

Reference /6/ describes three different types of techniques for a frequency management:

I Predictions a) long term
b) short term
II Soundings
III Channel Evaluation

Long term prediction can be used to provide data for system planning, frequency assignment to groups of users and the selection of frequencies with its seasonal, monthly and daily changes. With sophisticated propagation analysis programs /3/ the prediction of different path parameters - operational frequencies (MUF, maximum usable frequency, LUF, lowest usable frequency), path attenuation, signal to noise ratio etc. - is possible with reasonable reliability. Long term predictions must cover a complete sunspot cycle.

Short term prediction involve measurements of currently prevailing inospheric conditions to update the long term predictions. With these update it is possible to take care of any anomalous propagation mechanism that might be present. Daily forecast of typical parameters (sun spot numbers, critical frequency) are available from different "radio observing centers" /10/.

With sounding methods it is possible to measure propagation characteristics such as channel impulse response, signal propagation delay and signal amplitude.

Channel evaluation provides information on the actual transmission like signal to noise ratio and data error rate, respectively.

The three techniques listed above are different in complexity and effectiveness. The implementation in a frequency management systems depends on the realisation effort.

The accurancy of long term prediction programs, supported by a permanent up date of actual ionospheric parameters, is sufficient for most applications. Therefore they play an important role in todays frequency management systems /6,8/. With off-line propagation analysis procedures it is possible to provide frequency selection data, which will guarantee a high probability of a satisfactory link quality at any time. Typically, the useful frequency window in the MUF-LUF range is predicted on an hour by hour basis to find an optimum working frequency (OWF) see (Fig. 4-1, MUF-LUF range). With this off-line method it is also possible to predict the expected SNR with relative accurancy if the necessary system parameters are given. Long term prediction programs like IONCAP are available for use on PCs and they can also be integrated into special radio link processors.

Despite of continuously refinements, propagation analysis programs have certain fundamental limitations. One of the main limitations is, that effects of additve interference coming from other frequency band users are not included in the analysis and prediction model. To overcome the limitations in long-term forecasting techniques, real-time channel evaluation methods have been developed.

The second method to analyse the propagation medium is ionospheric sounding. The aim of inospheric sounding is to provide propagation characteristics such as channel unit impulse response, signal propagation delay and signal amplitude. Figure 4-2 shows a typical picture of an ionogram.

In conventional HF communication systems data transmission methods are used which cause intersymbol interference in case of multipath fading. With sounding measurements a region of single mode propagation can be selected to avoid this effect. This selection is usually done by a radio operator and because of the dynamic changes of the ionospheric reflection this is without success in many cases.

Modern HF communication system use serial modems with channel equalization which use the multipath propagation in a positive sense (so-called "implicit diversity"), only the SNR is important.

The system concept, described in the previous chapter, contains a transmission method with channel equalization. In this case the frequency management concept does not require sounding strategies with respect to multipath propagation, i. e. channels with multipath can be used for transmission as well.

The third method to determine the actual propagation condition is the real time channel evaluation (RTCE). In /7/ a definition of RTCE describes that the main task consists in measuring appropriate parameters of the channel in real-time and to derive of a numerical transmission model for each individual channel. With this model a qualitiy of each channel can be derived and this can be used for frequency selection.

Many different forms of RTCE systems have been developed which use a variety of parameters. One of the generally used parameters is the noise or interference level. This is particulary important, because in many cases interference is the limiting factor of communication performance, rather than propagation effects.

The realization of an adaptive frequency management concept, based on an RTCE-procedure which involve measurements of interference and bit error rate (BER) will be described in the next section.

4.2 Realization of an adaptive frequency management system

The adaptive frequency management system of the radio link processor described in section 3.3 performs different functions, active and passive, to provide the best frequency (OWF, optimum working frequency) at a given time. The criteria for the frequency selection is the circuit reliability CR. The CR is defined in /8/ as the probability, that the expected SNR exceeds a required level, multiplied by the probability that the sky-wave path for the specified frequency exist, i. e.:

$$CR = q_f q_{S/N} \tag{4-7}$$

q_f = fraction of days, during which a sky-wave path of a specified frequency is expected to be present

q_{SNR} = probability, that the actual signal-to noise ratio will exceed the required SNR.

The actions within the realized frequency management consist of three steps which determine the CR as described before. Figur 4-3 shows the basic functions of the frequency management with its three steps of channel assessment. A fundamental task in this assessment is, to distinguish how fast the individual interfering influence change on the channels. The reaction times of the countermeasures have to be adapted according to these typical interference time pattern.

In the first step a frequency sub-pool (SP) is selected from a main frequency pool (MP) by a hourly adaptation. The actual limits of the usable frequency range are determined, as a function of the hour of day,the sunspot number, season, geographical location, distance. The adaptation of the frequency sub-pool to the progress of the day, can be archieved with relatively high accuracy using long-term propagation prediction charts /3/. The result of the first step is a determination of a frequency window according to the probability q_f that the sky-wave mode is present. The frequencies are sorted in the subpool table according to their suitability for a transmission as a function of FOT-MUF-HPF (FOT, optimum traffic frequency, HPF, highest possible frequency) progress.

In step two the frequencies of the sub-pool currently being used are continuously checked by a channel monitoring facility to detect actual interference and accupancy by other users. The usefulness of the pool frequencies is examined by a measurement of noise level. The result is entered into a channel monitoring table CM. In this table the frequencies are sorted according to their lowest interference level and their distance to the FOT.

In the third step an active channel analysis will be used to estimate the link quality (q_{link}) of the operating frequency. After establishing the link, the transmission quality is continuously monitored. With measurements of ARQ repetitions a link statistic table (LS) is contructed. The result is an estimation of the actual value for q_{link} for situations associated with different link distances.

With these three steps as show in figure 4-3 the selection of the optimum working frequency OWF is possible. With the combination of the three steps a modified circuit reliability $\widetilde{CR}$ can be defined (4-8)

$$\widetilde{CR}(f_i) = q_{FOT}(f_i)\, q_{link}(f_i) \qquad i = 1,20 \tag{4-8}$$

$q_{FOT}(f_i)$ = probability, that a sky-wave path for the frequency f_i exists with the respect to the actual FOT

$q_{link}(f_i)$ = probability, that the transmission quality of the frequency f_i will exeed a required quality (no. of ARQ repetitions)

The criteria for the OWF is the maximum value of $\widetilde{CR}$:

$$\max_{i = 1,20} \{\widetilde{CR}(f_i)\} = \widetilde{CR}(f_{OWF}) \tag{4-9}$$

In reference /2/ the radio link processor is described in which this adaptive frequency management is realized.

In the next section some results of field test with this adaptive frequency management are described.

4.3 Results of field test

In order to evaluate the performance of the implemented frequency management extensive field tests with a communication system were carried out. The main feature, the automatic frequency selection procedure, was tested on different link distances. Trials were made in Germany with distances of 40 km, 80 km and 120 km. In each case only the skywave was present. Some typical results of these trials are described in the following.

In order to obtain results which shows the performance of the frequency selecting procedure, different measurements of internal quantities of the adaptive frequency management system were made during the trials. The following pictures show four different typs of results.

Figure 4-4.1 (distance 40 km) shows the interference level histogram of a specified frequency sub-pool (20 frequencies) versus time of day. These noise levels were measured with the channel monitoring facility, which is a part of the adaptive

frequency management system. As it is know for long time the noise level at day time is lower than at night time. As a result of many field tests, an average noise level of -100 dBm can be determined to get a succesful transmission.

Figures 4-5.1 and 4-6.1 show nearly the same noise enviroment during the test of other link distances and different days of measurements.

As a result of Figure 4-4.1, the number of frequencies available for good transmission (taken from the hourly adapted subpool), is shown in figure 4-4.2, 4-5.2 and 4-6.2. For the trials with 40 km and 80 km distances the sup-pool consist of 20 frequencies at every time, for the trials of 120 km it contains only 6 frequencies. From this figure it can be concluded that a relative high availability at day time resulted. This is in contrast to night time were the high interference level reduces the availability of useful frequencies.

To measure the performance of the frequency selection procedure the trials were support by an automatic procedure, which starts a link set up every five minutes. During a period of 24 hours 288 link set ups were possible. The communication system has a threshold with a maximum of 10 calls to establish the link. After 10 calls the link set up procedure is stopped automatically.

These figures show the number of calls to establish a link versus time of day. The result shows a good performance of link set up at day time. At night time the number of calls increases because of the lower availability of useful frequencies. The result of the link test, where only 6 frequencies are used, shows this effect in figure 4-6.4 very clear. During the morning time no link set up was possible.

Figure 4-4.4, 4-5.4 and 4-6.4 show the result of link set up measurements. They show the precentage of link set up at day and night versus the number of calls. At day time a average of 1.5 calls per link set up results and at night time a average of 1.8 calls. In case of these trials, where only 6 frequencies were used for day and night time, the average number of calls at night time increases to 6.75. The reason is the very low of availability of useful frequencies at night as shown in figure 4-6.2.

The performance of the adaptive frequency management can be seen in more detail in the figure 4-4.3, 4-5.3 and 4-6.3.

In conclusion, the examples of different link test show the effectivenes of the frequency management algorithms. The procedure of selecting the optimum working frequency from a hourly adapted sup-pool leads to a good link performance and circuit reliability. The essential result of the field tests was that the sub-pool must contain useful frequencies at any time.

5. CONCLUSION

The transmission medium HF radio has properties, which have a strong influence on the design of modern HF data communication systems. It has been shown, that only such a system approach can give real improvements compared with conventional HF transmission.

An HF modem is necessary within the system, which occupies enough bandwidth to cope with multipath propagation (or frequency selective fading) and which has adaptation algorithms, which track also fast variations of multipath. Furthermore the modem must cope with strong interference or jamming. FEC and ARQ are necessary and a robust protocol which handles every situation within the radio network and on the transmission medium. The protocol's control information must be protected with more redundancy to assure fast reaction times also in case of strong interference. The protocol should also automatize the entire transmission of data packets within the network and it must leave only those activities on the user side, which can not be done by the system automatically. It must perform link set up and changing of frequencies during transmission if necessary.

To do all this in the best way it must be supported by an internal adaptive frequency management, which manages the allocated bundel of frequencies dynamically. It must observe some external input quantities coming from the user, e. g. time of day, month, distances of radio stations, and some internal quantities resulting from measurements, e. g. interference levels or bit error rates. It has to learn the quality of the available frequencies and its dependence on time.

This paper is based on a realization of an HF data communication network, in which all the aspects discussed have been implemented. Simulations and results of tests with radio stations on real HF channels show that the system works as it is intended.

6. REFERENCES

/1/ HF Radio Systems and Techniques, IEE Conference, London, Conf. Publication Nr. 245, 1988

/2/ Borgmann D.: An HF-Radio Link Processor with Adaptive Frequency Management. IEE Conf. Publication Nr. 245, 1988

/3/ Teters L.R.; Lloyd J.L.: Ionospheric Communication Analysis and Prediction Program. IONCAP, NTIA Report 83-127, July 1983

/4/ Watterson, C.C. et al: An Ionospheric Channel Simulator. ESSA Technical Memorandum ERLTM - ITS 198, US-Department of Commerce, 1969

/5/ Peterson, W.W.; Weldon, E.J.: Error Correcting Codes. Second Edition, MIT Press, 1981

/6/ Nicholas M. Maslin: HF Communication: A System Approach. Pitman Publishing, 1987

/7/ Dawson, M. Darnell: An HF-System Design with Embedded Channel Evaluation. IEE Conf. Publication Nr. 245, 1988

/8/ Charles M. Rush: Ionospheric Radio Propagation Model and Predictions - A Mini Review. IEEE Vol. AP-34, No. 9, Sept. 1986

/9/ Darnell M.: Real Time Channel Evaluation. AGARD LECTURES Series No. 127, 1983

/10/ URSI, International URSIOGRAM and World Days Service

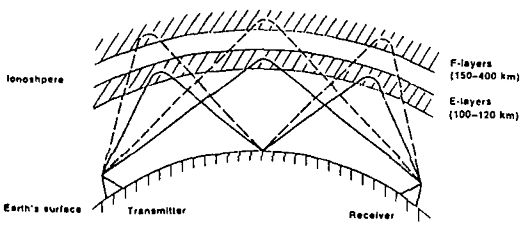


Fig. 2-1 Multipath propagation on shortwave radio

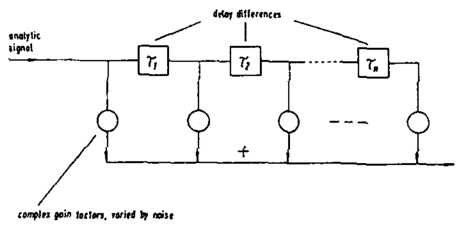


Fig. 2-3 Model of multipath propagation

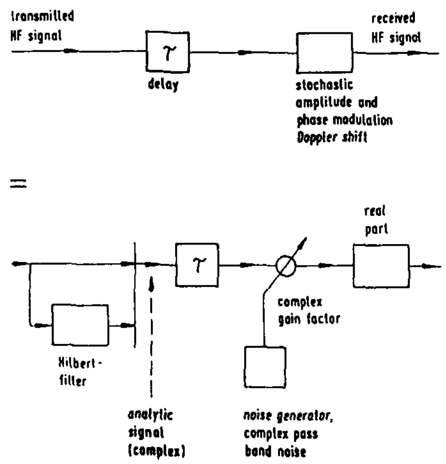


Fig. 2-2 Model of single fading path

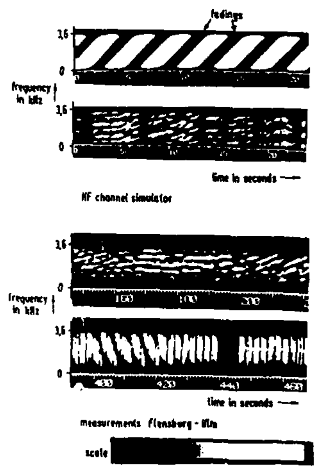
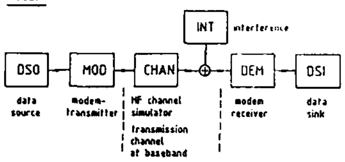


Fig. 2-4 Fadeograms

Model:



Parameters:

- HF channel simulator: • number of paths • delay • doppler shift • doppler bandwidth • relative amplitude (rms) } for each path
- Interference: • white Gaussian noise • atmospheric impulse noise • CW interference • FSK interference • ASK interference • PSK interference

Fig. 2-5 Model of the HF radio channel

ISO/OSI layer

7	application	application specific layer
6	presentation	
5	session	
4	transport	
3	network	network/transmission specific layers
2	data link	
1	physical	

Fig. 3-1 Layers of the ISO/OSI model

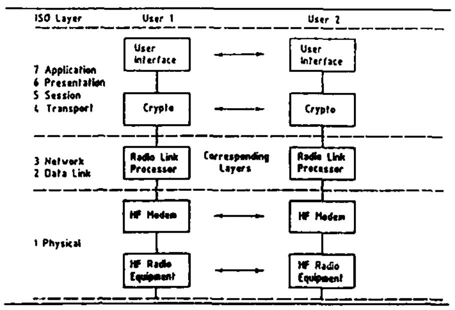
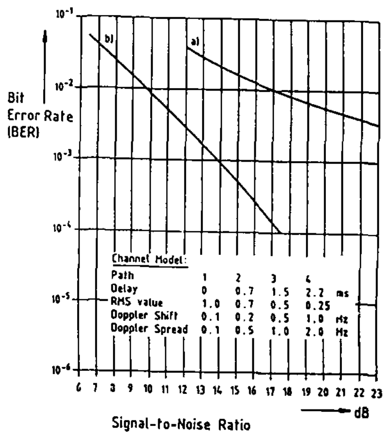


Fig. 3-2 HF radio communication architecture model

Operation:	duplex
Modulation:	single tone 4 phase shift keying (QPSK)
Carrier frequency:	1650 Hz
Data rate:	9600,4800 bps (intermittent) 2400-50 bps (continuous)
Bandwidth:	max. 2.7 kHz (300 Hz - 3 kHz)
Demodulation:	coherent; DFE, continuously adaptated
Acquisition time:	70 ms
Burst signal transmission:	70 ms + data
Acceptable multipath spread:	5 ms
Frequency offset correction:	± 75 Hz initial value, tracking up to 3.5 Hz/s
Error correction:	FEC with concatenated codes (Reed-Solomon+bi-orthogonal)
Interleaving:	2s, 10s, 100s

Fig. 3-3 ECHOTEL features



a) parallel tone modem
b) ECHOTEL modem

Fig. 3-4 Comparison of two modems with respect to bit error rate, without FEC

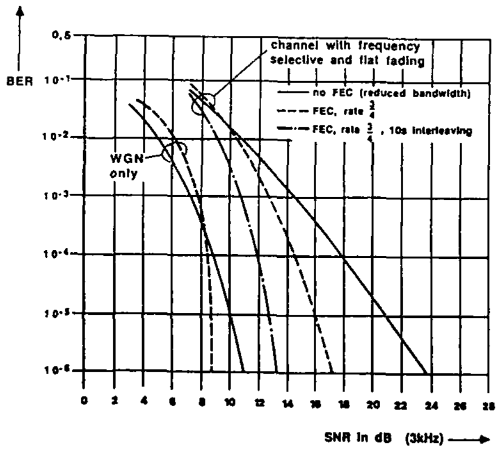
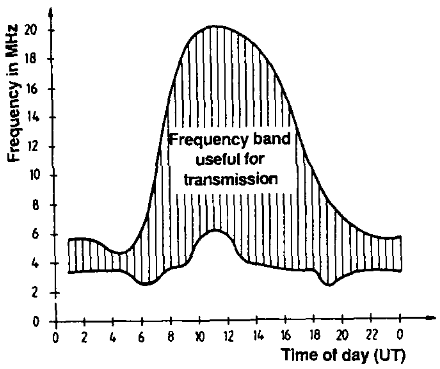


Fig. 3-5 ECHOTEL modem 2400 bps
Influence of FEC and interleaving on BER for different conditions on the channel



Link:	**Flensburg-Ulm**
Month:	**December**
Transmitted power:	**500W**
Transmitting antenna:	**Horizontaldipol**
Receiving antenna:	**Horizontaldipol**
Sun spot number:	**200**
required reliability:	**0,9**
required SNR/Bandwitdh:	**45dB/Hz**

Fig. 4-1 **Frequency window useful for transmission and the various parameters of dependence**

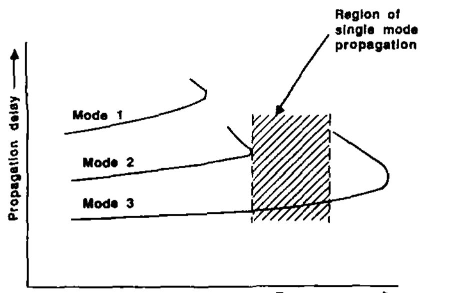


Fig. 4-2 **Ionogram with region of single mode propagation**

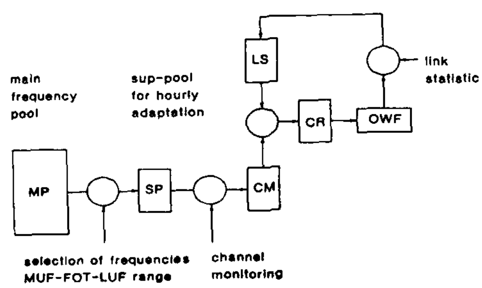


Fig. 4-3 **Basic functions of frequency management**

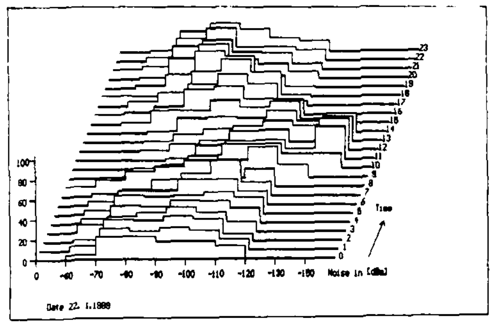


Fig. 4-4.1 **Link Test: distance 40 km**
Noise histogram versus time of day

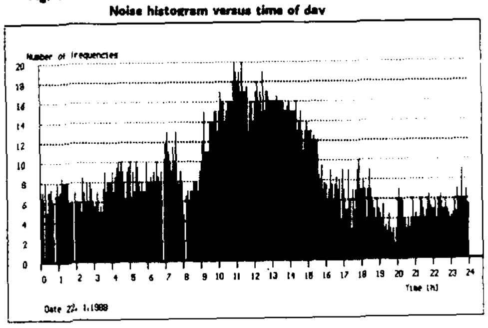


Fig. 4-4.2 **Link Test: distance 40 km**
Number of available sub-pool frequencies versus time of day

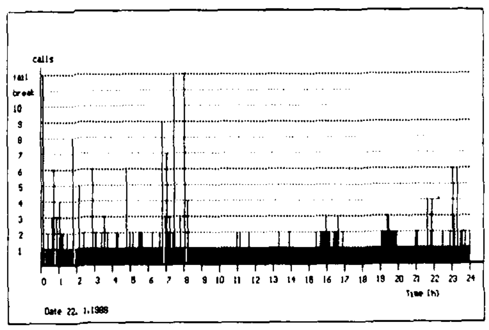


Fig. 4-4.3 **Link Test: distance 40 km**
Number of calls per link set up versus time of day

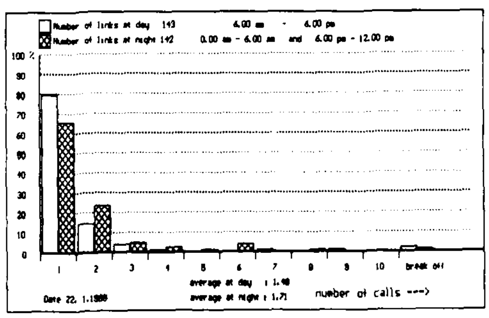


Fig. 4-4.4 **Link Test: distance 40 km**
Procentage of link set up at day/at night versus number of calls

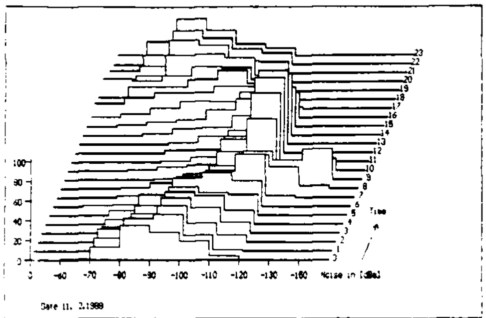


Fig. 4-5.1 Link Test: distance 80 km
Noise histogram versus time of day

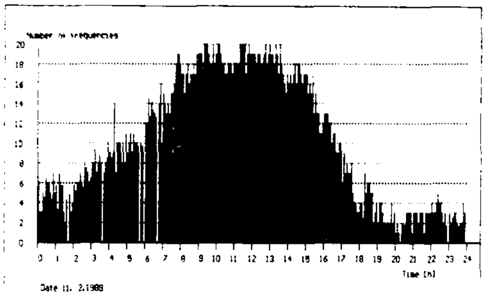


Fig. 4-5.2 Link Test: distance 80 km
Number of available sub-pool frequencies versus time of day

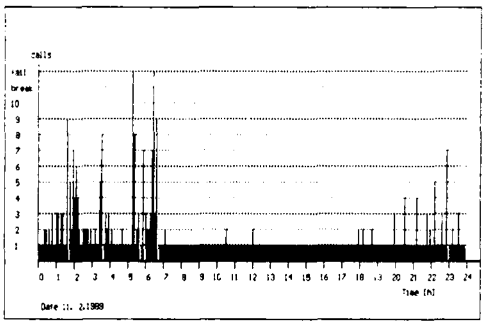


Fig. 4-5.3 Link Test: distance 80 km
Number of calls per link set up versus time of day

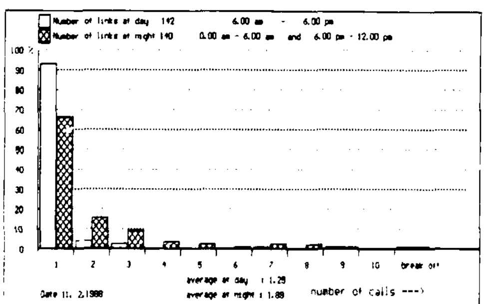


Fig. 4-5.4 Link Test: distance 80 km
Precentage of link set up at day/at night versus number of calls

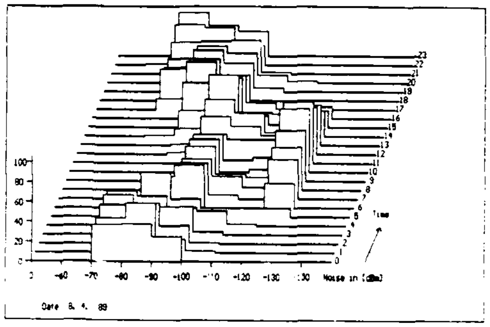


Fig. 4-6.1 Link Test: distance 120 km
Noise histogram versus time of day

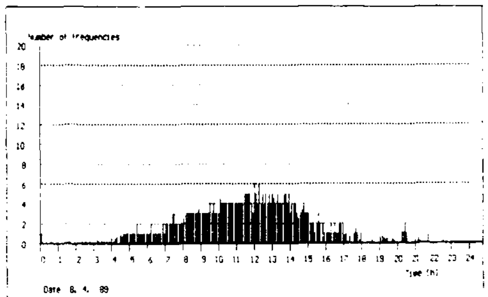


Fig. 4-6.2 Link Test: distance 120 km
Number of available sub-pool frequencies versus time of day

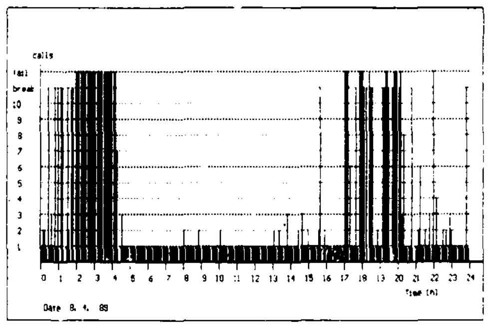


Fig. 4-6.3 Link Test: distance 120 km
Number of calls per link set up versus time of day

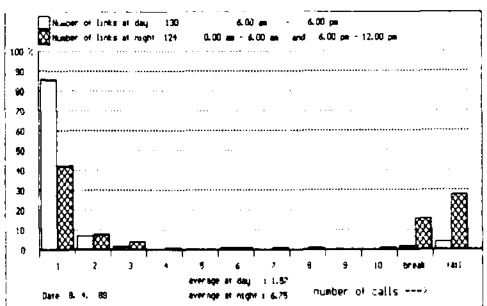


Fig. 4-6.4 Link Test: distance 120 km
Precentage of link set up at day/at night versus number of calls

LIST OF PARTICIPANTS

ALBRECHT, H.J. Dr	FGAN, Neuenahrerstrasse 20, D-5307 Wachtberg-Werthhoven, Federal Republic of Germany
ALLEN, K.C. Mr	ITS-S3, 325 Broadway, Boulder, CO 80303, United States
Commander and Director	US Army Engineering Topographic Laboratories, Attn: CEETL-GL-AT (Mr J E.Allen), Fort Belvoir, VA 22060-5546, United States
ALTAY, T. Maj.	3ncu Hava İkmal ve Bakım Merkezi K.lığı, Etimesgut, Ankara, Turkey
ANDERSON, K.D. Mr	Naval Ocean Systems Center, Code 543, San Diego, CA 92152-5000, United States
ANTE, E. Mr	ET-K, ESG, Elektronik-System GmbH, Vogelweideplatz 9, D-8000 München 80, Federal Republic of Germany
AUDONE, B. Ing.	Aeritalia, Avionics Systems and Equipment Gp, 10072 Caselle Torinese, Italy
BARRIOS, A.E. Ms	Naval Ocean Systems Center, Code 543, San Diego, CA 92152-5000, United States
BAUMGARTNER, G.B. Dr	Naval Ocean Systems Center, Code 543, San Diego, CA 92152-5000, United States
BELROSE, J.S. Dr	PO Box 11490, Communications Research Center, Station H, Ottawa, ON K2H 8S2, Canada
BENSON, R.F. Dr	Interplanetary Physics Branch, Lab. for Extraterrestrial Physics, Space and Earth Sciences Dir., NASA Goddard Sp. Fl. Cter/Code 692, Greenbelt, MD 20771, United States
BERNALDO DE QUIROS, J.G. Dr	E.T.S.I., Aeronauticos, Plaza Cardenal Cisneros, 3, 28040 Madrid, Spain
BISHOP, G. Mr	AFGL/LIS, Hanscom AFB, MA 01731-5000, United States
BISSONNETTE, L.R. Dr	Defence Establishment Valcartier, 2459 Pie XI Blvd North, (PO Box 8800), Courcelette, Quebec, G0A 1R0, Canada
BROOKER, R.W. Cdr.	SPAWARSYSCOM (3211), National Center #1, Washington, DC 20363-5100, United States
BORGMANN, D. Dipl. Ing.	AEG, Geschaftsbereit Hochfrequenz-Technik, Postfach 1730, Sedanstrasse 10, D-7900 Ulm/Donau, Federal Republic of Germany
BOSSY, L. Prof.	174, Av. Winston Churchill, UCCLE, B-1180 Bruxelles, Belgium
CANN, G.A. Mr	General Dynamics Corporation, DIO/UWC Complex, 1525 Wilson Blvd, Suite 700, Arlington, VA 22209, United States
CANNON, P.S. Dr	Flight Management, P 161 Building, Royal Aerospace Establishment, Farnborough, Hants GU14 6TD, United Kingdom
CARATORI, J. Dr	LETTI, Université Paris-Sud, 9, Av. de la Division Leclerc, 94230 Cachan, France
CHRISTOPHE, F. Mr	Dept. Micro-Ondes, ONERA-CERT Toulouse, BP 4025, 2, Av E. Belin, 31055 Toulouse Cedex, France
CLIFFORD, S.F. Dr	Director of Wave Propagation Lab., National Oceanic and Atmospheric Administration NOAA/ERL/WPL-R/E/WP, 325, Broadway, Boulder, CO 80303, United States
COOK, W.A. Lt Cdr.	Spac and Naval Warfare Systems Command, Washington, DC 20363-5100, United States
COTE, O.R. Mr	EOARD, Box 14, E.P.O. New York 09510-0200, United States
CRAIG, K.H. Dr	Bld R25, Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QW, United Kingdom
CRESS, T.S. Col.	OSUD (A) ODDR&E (R&AT) MET, The Pentagon, Room 3D129, Washington, DC 20301-3080, United States
DAMBOLDT, T. Dr	FI der DBP, PO Box 5000, D-6100 Darmstadt, Federal Republic of Germany
DANIEL, A. Dr	INSA — LCST, 20, Av. des Buttes de Coesmes, 35043 Rennes Cedex, France
DAVIDSON, K.L. Prof.	Naval Postgraduate School, Dept. of Meteorology, 63 DS, Monterey, CA 93940, United States

DAVIDSON, J.D., Jr Capt.	HQ/AFOTEC/WE, Kirtland AFB, NM 87117-7001, United States
DOCKERY, G.D. Mr	JHU/Applied Physics Laboratory, Johns Hopkins Road, Laurel, MD 21044, United States
DOS SANTOS, A.C. Capt.	CLAFA, Dir. de Electrotecnica, Base de Alfragide, 2700 Amadora, Portugal
DURAND, J-C. Mr	SDC/DT/ETT, 18, Av. du Maréchal Juin, 92363 Meudon la Forêt, France
DURANDO, G.B. Lt	Mariteleradar, Viale Italia 72, 57100 Livorno, Italy
EL ZEIN, G. Ph. Dr	INSA — LCST, 20, Av. des Buttes de Coesmes, 35043 Rennes Cedex, France
FANG, D.J. Mr	Massachusetts Technological Laboratory, Inc., MTL Washington Division, 4300 Montgomery Avenue, STE 204, Bethesda, MD 20814-4402, United States
FELSEN, L.B. Prof.	Dept of Electrical Engineering and Computer Science, Polytechnic University, Farmingdale, NY 11735, United States
FERGUSON, J.A. Dr	Naval Ocean Systems Center, Code 542(T), San Diego, CA 92152-5000, United States
FITZSIMONS, T.K. Dr	Allied Radio Frequency Agency, NATO HQs, 1110 Bruxelles, Belgium
FLOOD, W.A. Dr	Director Geosciences Division, US Army Research Office, PO Box 12211, Research Triangle Park, NC, 27709-2211, United States
FOURNIER, M. Dr	LCTAR, 6, rue Nieuport, BP 16, 78143 Vélizy-Villacoublay, France
FRANCALACCI, F. Cdr.	Mariteleradar, Viale Italia, 72, 57100 Livorno, Italy
GOLD, J.M. Mr	British Aerospace Ltd (Dynamics), Bristol, FPC-213, PO Box 5, Filton, Bristol, BS12 7QW, United Kingdom
GOMEZ, R.B. Dr	US Army Engineer Topographic Lab., Attn: CEETL-TD-A, Fort Belvoir, VA 22060-5546, United States
GONCALVES, C.A. Mr	Instituto Nacional de Meteorologia e Geofisica, Rue C do Aeroporto, 1700 Lisboa, Portugal
GOODMAN, J.M. Dr	Naval Research Laboratory, Code 4180, Washington, DC 20375-5000, United States
GORMAN, F. Mr	US Army Communications Electronics Command, AMSEL-RD-C3-TR-1, Fort Monmouth, NJ 07703, United States
GOUTELARD, C. Prof.	Directeur L.E.T.T.I., Université Paris-Sud, 9, Av. de la Division Leclerc, 94230 Cachan, France
HAGN, G.H. Mr	Assistant Director, Information and Telecom. Sciences Center, SRI International, 1611 N.Kent Street, Arlington, VA 22209-2173, United States
HAINES, M. Mr	University of Lowell, Center for Atmospheric Research, 450 Aiken Street, Lowell, MA 01854, United States
HANCOCK, J. Major	AFGL/LYA, Hanscom AFB, MA 01731, United States
HARRISON, E.J., Jr Dr	COMSPANAVWAYSSYSCOM, PMW 141 — Room 3E92, Washington DC 20363, United States
HATTAN, C.P. Mr	Naval Ocean Systems Center, Code 543, San Diego, CA 92152-5000, United States
Commander and Director	US Army Atmospheric Sciences Laboratory, Attn: SLCAS-AE-O (Dr Heaps), WSMR, NM 88002-5501, United States
HELVEY, R. Mr	Head, Atmospheric Sciences Section, Geophysics Sciences Branch, Pacific Missile Test Center, Point Mugu, CA 93042, United States
HINKELMANN, G.E. Mr	Industrieanlagen Betriebsges., Einsteinstrasse, 8012 Ottobrunn, Federal Republic of Germany
HITNEY, H.V. Mr	Naval Ocean Systems Center, Code 543 (S), San Diego, CA 92152-5000, United States
HITNEY, L.R. Ms	Naval Ocean Systems Center, Code 542, San Diego, CA 92152-5000, United States
HÖHN, D. Dr	Direktor, Forschungsinstitut für Optik (FfO), Schloss Kressbach, 7400 Tübingen, Federal Republic of Germany
HOOGESTEYN, O. Ir	Ministerie van Defensie, DMKC/MBA-I/CvP, Van den Burchlaan 31, 2597 PC Den Haag, Netherlands
HUGHES, H.G. Mr	Naval Ocean Systems Center, Code 543(S), San Diego, CA 92152-5000, United States

JENSEN, J.J. Capt. USN	COMSPAWARSYSCOM (PMW-141), Washington, DC 20363-5100, United States
KEIZER, W.P.M.N. Mr	FEL-TNO, PO Box 96864, 2597 AK The Hague, Netherlands
KEMP, A.H. Mr	Dept of Electronic Engineering, University of Hull, Hull HU6 7RX, United Kingdom
KRAISS, K.F. Dr	FGAN/FAT, Neuenahrerstrasse 20, D-5307 Wachtberg-Werthhoven, Federal Republic of Germany
KRIEZIS, E. Prof.	Dept. of Electrical Engineering, Faculty of Technology, Aristotle Univ. of Thessaloniki, GR 54006 Thessaloniki, Greece
KUGEL, C.P. Mr	Naval Ocean Systems Center, Code 542, San Diego, CA 92512-5000, United States
KUSCHEL, H. Mr	FGAN-FHP, Neuenahrerstrasse 20, D-5307 Wachtberg-Werthhoven, Federal Republic of Germany
LANGOT, L. Mr	DRET, 26, Bd Victor, 75015, Paris, France
LARSEN, R. Mr	GEC Marconi Research Centre, Propagation Group, Great Baddow, Chelmsford, Essex CM2 8HN, United Kingdom
LASSUDRIE-DUCHESNE, P. Dr	CNET—LAB-MER-GER, Route de Trégastel, BP 40, 22301 Lannion Cedex, France
LEEUW, G. de Dr	Physics and Electronics Laboratory TNO, PO Box 96864, 2509 JG The Hague, Netherlands
LEONG, S.H. Dr	Naval Surface Warfare Center, Code F 42, Dahlgren, VA 22405, United States
LEVY, M.F. Dr	Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, United Kingdom
LINNARTZ, J.P. Dr	Telecommunication and Traffic Control Systems Group TVS, Fac. of El. Eng., 2013, Delft Univ. of Techn. PO Box 5031, 2600 GA Delft, Netherlands
MALAGA, A. Dr	Signatron, Inc., 110 Hartwell Avenue, Lexington, MA 02173-3198, United States
MICHELSON, R.L. Mr	Dept of Electrical & Computer Eng., University of Illinois, 1406, W.Green Street, Urbana, IL 61801, United States
MILLER, C.R. Mr	ORI, Inc., 6295 A Edsall, Rd, Alexandria, VA 22312-2617, United States
MOISION, W.K. Mr	Naval Ocean Systems Center, Code 542, San Diego, CA 92152-5000, United States
PAPPERT, R.A. Dr	Naval Ocean Systems Center, Code 542(T), San Diego, CA 92152-5000, United States
PARL, S.A. Dr	Signatron, Inc., 110 Hartwell Avenue, Lexington, MA 02173-3198, United States
PATTERSON, W. Mr	Naval Ocean Systems Center, Code 543, San Diego, CA 92124-5000, United States
PAUL, A.K. Dr	Naval Ocean Systems Center, Code 542(T), San Diego, CA 92152-5000, United States
PAULUS, R.A. Mr	Naval Ocean Systems Center, Code 543(S), San Diego, CA 92152-5000, United States
PERRY, J.E. Mrs	1921 Hopefield Rd, Silver Spring, MD 20904, United States
POIRIER, J.L. Dr	Chief Applied Electromagnetics Div., RADC (EEC), Hanscom Air Force Base, MA 01731, United States
RASMUSSEN, J.E. Mr	Chief, Ionospheric Interactions Branch, AF Geophysics Laboratory, Hanscom Air Force Base, MA 01731, United States
RICHTER, J.H. Dr	Head, Ocean and Atmospheric Sciences Division, Code 54, Naval Ocean Systems Center, San Diego, CA 92152-5000, United States
RILEY, N.G. Mr	Dept of Electronic Engineering, University of Hull, Hull HU6 7 [illegible] d Kingdom
ROBERT, G.P. Mme	Chef de Gpe Etudes Hyperfréquences, CESDA, BP 77, St Mandrier, 83800 Toulon Naval, France
ROSE, R.B. Mr	Naval Ocean Systems Center, Code 772(S), San Diego, CA 92152-5000, United States
ROSENTHAL, J. Mr	Head, Geoph. Sciences Branch, Code 3253, Pacific Missile Test Center, Point Mugu, CA 93042, United States

ROSSLER, S. Mrs	ESG Elektronik-System GmbH, Vogelweideplatz 9, D-8000 München 80, Federal Republic of Germany
ROY, T.N. Mr	Naval Ocean System Center, Code 542, San Diego, CA 92152-5000, United States
RUELLE, N. Mr	CNET-LAB-MER-GER, Route de Trégastel, BP 40, 22301, Lannion, France
RYAN, F.J. Mr	Naval Ocean Systems Center, Code 541(B), San Diego, CA 92152-5000, United States
SAILORS, D.B. Mr	Naval Ocean Systems Center, Code 542, San Diego, CA 92152-5000, United States
SANJURJO NAVARRO, R. Dr	Aeronauticos, Plaza Cardenal Cisneros, 3, 28040 Madrid, Spain
SCHIMPF, H. Dr	FGAN/FHP-GA, Neuenahrerstrasse 20, D-5307 Wachtberg-Werthhoven, Federal Republic of Germany
SCHOLZ, D. Lt Col.	Director, NARFA GE, Fernmeldeamt der BW, Tomburgkaserne, 5308 Rheinbach, Federal Republic of Germany
SEGURA, A. Ing	INTA, Esteban Terredas, Departamento de Avionica, Carretera de Ajalvir, S/N, Torrejon de Ardoz, Madrid, Spain
SHELLMAN, C.H. Mr	Naval Ocean Systems Center, Code 542(T), San Diego, CA 92152-5000, United States
SITTROP, H. Mr	FEL-TNO, PO Box 96864, 2509 JG The Hague, Netherlands
SKRIVANEK, R.A. Mr	Geophysics Laboratory/LI, Hanscom AFB, Bedford, MA 01731, United States
SNYDER, F.P. Dr	C/o Dr Jerry A. Ferguson, Code 524, Naval Ocean Systems Center, San Diego, CA, 92152-5000, United States
SOICHER, H. Dr	US Army Communications-Electronics Command AMSEL-RD-C3-TR-4, Fort Monmouth, NJ 07703-5202, United States
SONLU, Y. Major	3 HIBM.K.Fb.Md. lüğü, 06375 Etimesgut, Ankara, Turkey
SPRENKELS, C. Col. d'Aviation Ir.	Etat-Major de la Force Aerienne, Sect. Comm. et Electronique (VDM), Quartier Reine Elisabeth, Rue d'Evère, B-1140 Bruxelles, Belgium
STAPLETON, J.K. Mrs	Naval Surface Warfare Center, Code F 42, Dahlgren, VA 22448-5000, United States
STROEMSNES, O. Mr	Norwegian Navy Material Command, N-5078 Haakonsvern, Bergen, Norway
TAAGHOLT, J. Mr	Danish Polar Center, 3, Hausergade, DK-1128 Copenhagen K, Denmark
TACCONI, G. Prof.	Universita'di Genova, Dipartimento Ingegneria Biofisica, Via all'Opera Pia 11/A, 16145 Genova, Italy
TAG, P.M. Dr	Naval Environmental Prediction Research Facility, Monterey, CA 93943-5006, United States
TAYLOR, L.L. Mr P.E.	The Mitre Corporation, Suite 201, 145 Wyckoff Road, Eatontown, NJ 07724, United States
TAYLOR, R.J. Dr	The Johns Hopkins University, Applied Physics Lab., Johns Hopkins Road, Laurel, MD 20707, United States
THEWS, E.R. Mr	Johns Hopkins University, Applied Physics Laboratory, Johns Hopkins Road, Laurel, MD 20707, United States
TWITCHELL, P.F. Dr	Office of Naval Research, Code 1214, 800 N. Quincy Street, Rm 528, Arlington, VA 22217-5000, United States
VIETH, R. Mr	Dornier System GmbH, PBWM, Postfach 1360, D-7990 Friedrichshafen 1, Federal Republic of Germany
VISSINGA, H. Ir.	Physics and Electronics Lab. TNO, PO Box 96864, 2509 JG The Hague, Netherlands
VLAHAKIS, Capt.	Hellenic Air Force, Research Technology Center KETA, Post Office Terpsitheas, 16501 Glyfada, Athens, Greece
WOLLENWEBER, F.G. Mr	Amt für Wehrgeophysik, W 14 Mont Royal, D-5580 Traben-Trarbach, Federal Republic of Germany
WOOLDRIDGE, F.R. Capt.	SPAWARSYSCOM (321), National Center # 1, Washington, DC 20363-5100, United States
YAVUS, D. Dr	Head Radio Branch, Communications Division, SHAPE Technical Center, PO Box 174, 2501 CD The Hague, Netherlands
YEH, K.C. Prof.	EL/University of Illinois, 1406, W.Green Street, Urbana, IL 6[illegible], United States

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's Reference	3. Further Reference	4. Security Classification of Document
	AGARD-CP453	ISBN 92-835-0524-7	UNCLASSIFIED

5. Originator Advisory Group for Aerospace Research and Development
North Atlantic Treaty Organization
7 rue Ancelle, 92200 Neuilly sur Seine, France

6. Title OPERATIONAL DECISION AIDS FOR EXPLOITING OR MITIGATING ELECTROMAGNETIC PROPAGATION EFFECTS

7. Presented at the Electromagnetic Wave Propagation Panel Symposium held in San Diego, California, USA, 15—19 May 1989.

8. Author(s)/Editor(s)	9. Date
Various	September 1989

10. Author's/Editor's Address	11. Pages
Various	492

12. Distribution Statement This document is distributed in accordance with AGARD policies and regulations, which are outlined on the Outside Back Covers of all AGARD publications.

13. Keywords/Descriptors

Decision aids (ODAs)	*Remote sensing*
Propagation anomalies	Environmental effects
Propagation modelling	Operational assessment

14. Abstract

Exploitation or mitigation of environmental effects rank equal in importance with weapons systems. The rapidly changing propagation environment throughout NATO's area of concern requires a rapid operational assessment capability.

Since operational decision aids require appropriate propagation models and accurate environmental inputs, attention must be paid to propagation modelling, direct and remote sensing techniques and environmental forecasting.

To this end, the meeting covered the following topics: General aspects, propagation modelling and validation, decision aids for tropospheric radio progation, decision aids for ionospheric radio propagation, decision aids for electro-optical propagation, remote and direct sensing techniques, forecasting of environmental parameters and operational systems impact.

AGARD Conference Proceedings No.453
Advisory Group for Aerospace Research and Development, NATO
OPERATIONAL DECISION AIDS FOR EXPLOITING OR MITIGATING ELECTROMAGNETIC PROPAGATION EFFECTS
Published September 1989
492 pages

Exploitation or mitigation of environmental effects rank equal in importance with weapons systems. The rapidly changing propagation environment throughout NATO's area of concern requires a rapid operational assessment capability.

Since operational decision aids require appropriate

P.T.O.

AGARD-CP-453

Decision aids (ODAs)
Propagation anomalies
Propagation modelling
Remote sensing
Environmental effects
Operational assessment

AGARD Conference Proceedings No.453
Advisory Group for Aerospace Research and Development, NATO
OPERATIONAL DECISION AIDS FOR EXPLOITING OR MITIGATING ELECTROMAGNETIC PROPAGATION EFFECTS
Published September 1989
492 pages

Exploitation or mitigation of environmental effects rank equal in importance with weapons systems. The rapidly changing propagation environment throughout NATO's area of concern requires a rapid operational assessment capability.

Since operational decision aids require appropriate

P.T.O.

AGARD-CP-453

Decision aids (ODAs)
Propagation anomalies
Propagation modelling
Remote sensing
Environmental effects
Operational assessment

AGARD Conference Proceedings No.453
Advisory Group for Aerospace Research and Development, NATO
OPERATIONAL DECISION AIDS FOR EXPLOITING OR MITIGATING ELECTROMAGNETIC PROPAGATION EFFECTS
Published September 1989
492 pages

Exploitation or mitigation of environmental effects rank equal in importance with weapons systems. The rapidly changing propagation environment throughout NATO's area of concern requires a rapid operational assessment capability.

Since operational decision aids require appropriate

P.T.O.

AGARD-CP-453

Decision aids (ODAs)
Propagation anomalies
Propagation modelling
Remote sensing
Environmental effects
Operational assessment

AGARD Conference Proceedings No.453
Advisory Group for Aerospace Research and Development, NATO
OPERATIONAL DECISION AIDS FOR EXPLOITING OR MITIGATING ELECTROMAGNETIC PROPAGATION EFFECTS
Published September 1989
492 pages

Exploitation or mitigation of environmental effects rank equal in importance with weapons systems. The rapidly changing propagation environment throughout NATO's area of concern requires a rapid operational assessment capability.

Since operational decision aids require appropriate

P.T.O.

AGARD-CP-453

Decision aids (ODAs)
Propagation anomalies
Propagation modelling
Remote sensing
Environmental effects
Operational assessment

propagation models and accurate environmental inputs, attention must be paid to propagation modelling, direct and remote sensing techniques and environmental forecasting.

To this end, the meeting covered the following topics: General aspects, propagation modelling and validation, decision aids for tropospheric radio progation, decision aids for ionospheric radio propagation, decision aids for electro-optical propagation, remote and direct sensing techniques, forecasting of environmental parameters and operational systems impact.

Papers presented at the Electromagnetic Wave Propagation Panel Symposium held in San Diego, California, USA, 15—19 May 1989.

ISBN 92-835-0524-7

propagation models and accurate environmental inputs, attention must be paid to propagation modelling, direct and remote sensing techniques and environmental forecasting.

To this end, the meeting covered the following topics: General aspects, propagation modelling and validation, decision aids for tropospheric radio progation, decision aids for ionospheric radio propagation, decision aids for electro-optical propagation, remote and direct sensing techniques, forecasting of environmental parameters and operational systems impact.

Papers presented at the Electromagnetic Wave Propagation Panel Symposium held in San Diego, California, USA, 15—19 May 1989.

ISBN 92-835-0524-7

propagation models and accurate environmental inputs, attention must be paid to propagation modelling, direct and remote sensing techniques and environmental forecasting.

To this end, the meeting covered the following topics: General aspects, propagation modelling and validation, decision aids for tropospheric radio progation, decision aids for ionospheric radio propagation, decision aids for electro-optical propagation, remote and direct sensing techniques, forecasting of environmental parameters and operational systems impact.

Papers presented at the Electromagnetic Wave Propagation Panel Symposium held in San Diego, California, USA, 15—19 May 1989.

ISBN 92-835-0524-7

propagation models and accurate environmental inputs, attention must be paid to propagation modelling, direct and remote sensing techniques and environmental forecasting.

To this end, the meeting covered the following topics: General aspects, propagation modelling and validation, decision aids for tropospheric radio progation, decision aids for ionospheric radio propagation, decision aids for electro-optical propagation, remote and direct sensing techniques, forecasting of environmental parameters and operational systems impact.

Papers presented at the Electromagnetic Wave Propagation Panel Symposium held in San Diego, California, USA, 15—19 May 1989.

ISBN 92-835-0524-7